# HYDRO-CLIMATIC FORECASTING

# USING SEA SURFACE TEMPERATURES

A Dissertation
Presented to
The Academic Faculty

by

Chia-Jeng Chen

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Civil and Environmental Engineering

Georgia Institute of Technology
August 2012

# HYDRO-CLIMATIC FORECASTING

# USING SEA SURFACE TEMPERATURES

Approved by:


Dr. Aris P. Georgakakos, Advisor
School of Civil and Environmental Engineering
*Georgia Institute of Technology*

Dr. Huaming Yao
School of Civil and Environmental Engineering
*Georgia Institute of Technology*

Dr. Terry W. Sturm
School of Civil and Environmental Engineering
*Georgia Institute of Technology*

Dr. Robert X. Black
School of Earth and Atmospheric Sciences
*Georgia Institute of Technology*

Dr. Jian Luo
School of Civil and Environmental Engineering
*Georgia Institute of Technology*


Date Approved: May 1, 2012

# DEDICATION

To my wife, Hsiao-Hui,
and in memory of our kindly grandfathers

# ACKNOWLEDGEMENTS

The completion of my dissertation would not have been possible without the support and nurture from my advisor, Dr. Aris Georgakakos. His guidance has not only influenced me to be meticulous in my research but also motivated myself to become a more tactful person.

I would also like to extend my gratitude to my committee members—Dr. Terry Sturm, Dr. Jian Luo, Dr. Huaming Yao, and Dr. Robert Black—for their time and insightful comments on this work.

I gratefully acknowledge assistance and encouragement I received from my current and previous colleagues—Dongha Kim, Feng Zhang, Ke Yu, Xi Chen, Martin Kistenmacher, Christian Braneon, Freddrick Kimaite, Jeffrey Regan, Raymond Kim, and Yannis Dialynas—at the Georgia Water Resources Institute. They never wavered in their friendship and moral support over these years.

I am deeply indebted to my wife, Hsiao-Hui, my parents and my sister for their unending love, consideration, and patience throughout my studying period. Their unparalleled support has given me the strength to confront any frustrations. I cannot thank them enough for what they have brought to my life.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF SYMBOLS

| | |
|---|---|
| $a_m$ and $b_m$ | canonical vectors |
| $Avg$ | spatial average operator |
| $B$ | the repetition number of a bootstrap experiment |
| **B** | matrix with canonical vector $b_m$ in each row |
| $corr$ | correlation operator |
| $D(r)$ | the odds ratio in the calculation of the GSS |
| $D(t)$ | dipole time series |
| $D_X$ and $D_Y$ | diagonal matrices that contain standard deviations of $X$ and $Y$ |
| **D** | the dipole predictor in linear regression |
| $e_m$ and $f_m$ | eigenvectors used to derive canonical vectors $a_m$ and $b_m$ |
| $f$ | forecasting value |
| $F$ | forecasting set |
| $GSS_{R,\alpha}$ | GSS threshold with a given significance level |
| $h$ | data sizes (number of years) |
| $I$ | the indicator function in the calculation of the Re |
| $J$ | the size of a contingency table |
| $K$ | the number of partitioned subsets in a CV procedure |
| $M$ | minimum dimension used in CCA |
| $Max$ | maximum operator |
| $Min$ | minimum operator |
| $n$ | the number of pixels in a predictor (SSTA) domain |

| | |
|---|---|
| $o$ | observation value |
| $[p \times q]$ | dipoles of size p and q pixels at each pole |
| $p_{ij}$ | the element of a performance matrix |
| **P** | performance matrix |
| $R^2$ | coefficient of determination |
| $R_C$ | diagonal matrix that contains canonical correlations |
| $s_{ij}$ | the element of a scoring matrix |
| **S** | scoring matrix |
| $v_m$ and $w_m$ | canonical variables |
| $X$ and $Y$ | data vectors used in CCA |
| **Y** | hydro-climatic predictand in linear regression |
| $\alpha$ | significance level |
| $\hat{\beta}$ | estimated linear regressive parameter |
| $\phi$ | the attribute of a pole |
| $\gamma$ | ensemble envelope width ratio |
| $\lambda$ and $\Lambda$ | Lagrange multipliers and function used in CCA |
| $\mu$ | statistical mean |
| $\rho$ | Pearson's correlation coefficient or canonical correlations |
| $\sigma$ | statistical standard deviation |
| $\Sigma$ | covariance matrix |
| $\xi$ | the number of selected dipoles used to generate ensemble forecasts |

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ACF | Apalachicola-Chattahoochee-Flint River Basin |
| AGCM | Atmospheric General Circulation Model |
| AMIP | Atmospheric Model Intercomparison Project |
| AMJ | April–June |
| AMJJASO | April–October |
| AMO | Atlantic Multidecadal Oscillation |
| CCA | Canonical Correlation Analysis |
| CGCM | Coupling Ocean-Atmosphere General Circulation Model |
| CLM | Climatological Forecasting |
| CMAP | CPC Merged Analysis of Precipitation |
| CPC | Climate Prediction Centers |
| CV | Cross Validation |
| DEMETER | Development of a European Multimodel Ensemble System for Seasonal to Interannual Prediction |
| DJF | December–February |
| DJFMA | December–April |
| DMI | Dipole Mode Index |
| DP | Dipole Method |
| DSP | Dynamical Seasonal Prediction |
| EA | East Atlantic |
| ECMWF | European Centre for Medium-Range Weather Forecasts |
| ENSO | El Niño-Southern Oscillation |
| EOF | Empirical Orthogonal Function |

| | |
|---|---|
| GSS | Gerrity Skill Score |
| HS | Hurricanes |
| IOD | Indian Ocean Dipole |
| IRI | International Research Institute |
| ITCZ | Inter-Tropical Convergence Zone |
| JAS | July–September |
| JFM | January–March |
| JJA | June–August |
| JJAS | June–September |
| LACF | Lower ACF |
| LDA | Linear Discriminant Analysis |
| LME | Lower-Minimum Excess (ensemble forecasting measure) |
| MAE | Mean Absolute Error |
| MAM | March–May |
| MEI | Multivariate ENSO Index |
| MLR | Multiple Linear Regression |
| MOHSST | Meteorological Office historical Sea Surface Temperature |
| NAO | North Atlantic Oscillation |
| NCAR | National Center for Atmospheric Research |
| NCEP | National Centers for Environmental Prediction |
| ND | November–December |
| NDJ | November–January |
| NDJFMA | November–April |
| NH | Northern Hemisphere |
| NN | Neural Network |

| | |
|---|---|
| NP | North Pacific |
| OI | Optimal Interpolation |
| ON | October–November |
| OND | October–December |
| ONI | Oceanic Nino Index |
| PDO | Pacific Decadal Oscillation |
| PNA | Pacific/North American |
| PRISM | Parameter-elevation Regressions on Independent Slopes Model |
| PROVOST | PRediction Of climate Variations On Seasonal to interannual Time scales |
| Re | Reliability (ensemble forecasting measure) |
| SLP | Sea Level Pressure |
| SO | Southern Oscillation |
| SON | September–November |
| SST | Sea Surface Temperature |
| SSTA | Sea Surface Temperature Anomaly |
| SVD | Singular Value Decomposition |
| TC | Tropical Cyclone |
| TEJ | Tropical Easterly Jet |
| TNH | Tropical/Northern Hemisphere |
| UACF | Upper ACF |
| UME | Upper-Maximum Excess (ensemble forecasting measure) |
| VIF | Variance Inflation Factor |
| WPO | West Pacific Oscillation |

# SUMMARY

A key determinant of atmospheric circulation patterns and regional climatic conditions is sea surface temperature (SST). This has been the motivation for the development of various teleconnection methods aiming to forecast hydro-climatic variables. Among such methods are linear projections based on teleconnection gross indices (such as the ENSO, IOD, and NAO) or leading empirical orthogonal functions (EOFs). However, these methods deteriorate drastically if the predefined indices or EOFs cannot account for climatic variability in the region of interest. This study introduces a new hydro-climatic forecasting method that identifies SST predictors in the form of dipole structures. An SST dipole that mimics major teleconnection patterns is defined as a function of average SST anomalies over two oceanic areas of appropriate sizes and geographic locations. The screening process of SST-dipole predictors is based on an optimization algorithm that sifts through all possible dipole configurations (with progressively refined data resolutions) and identifies dipoles with the strongest teleconnection to the external hydro-climatic series. The strength of the teleconnection is measured by the Gerrity Skill Score. The significant dipoles are cross-validated and used to generate ensemble hydro-climatic forecasts.

The dipole teleconnection method is applied to the forecasting of seasonal precipitation over the southeastern US and East Africa, and the forecasting of streamflow-related variables in the Yangtze and Congo Rivers. These studies show that the new method is indeed able to identify dipoles related to well-known patterns (e.g., ENSO and IOD) as

well as to quantify more prominent predictor-predictand relationships at different lead times. Furthermore, the dipole method compares favorably with existing statistical forecasting schemes. An operational forecasting framework to support better water resources management through coupling with detailed hydrologic and water resources models is also demonstrated.

# CHAPTER 1

# INTRODUCTION

An excessive swing in climate regimes can often lead to detrimental impacts on human society because of lack of preparedness. For instance, the 1997–1998 El Niño sprung devastating floods, droughts, forest fires, and crop failure in many world regions, burdening these regions with heavy reconstruction costs. A few years later, in August 2005, Hurricane Katrina ravaged Louisiana and other southeastern states, killing thousands of people and causing damage estimated at $200 billion. In 2007, the southeastern United States underwent the worst drought in over a century, exacerbating water sharing conflicts among Georgia, Alabama, and Florida and costing millions to the local economy.

Mitigating the adverse impacts of climate variability is possible with improvements in climate forecasting, characterizing upcoming climatic conditions month(s) or season(s) ahead. In practice, the distribution of forecasting products is the mission of such organizations as the National Centers for Environmental Prediction (NCEP), the Climate Prediction Centers (CPC), the International Research Institute (IRI), and the European Centre for Medium-Range Weather Forecasts (ECMWF). One of their goals is to develop dynamical models (Palmer et al., 2004; Saha et al., 2006) that couple atmospheric and oceanic interactions in a sophisticated manner. However, running dynamical models requires highly trained personnel, which is not usually available at

local agencies. In addition, dynamical model forecasts still lack the necessary skills and accuracy (Lavers et al., 2009).

Alternatively, statistical models that utilize macro-scale interconnections between oceanic and atmospheric variables have exhibited remarkable predictive value and operational utility and flexibility (Anderson et al., 1999; Gershunov and Cayan, 2003; Hastenrath, 1995). Thus, even though climate forecasting based on dynamical approaches is ideally preferable, scientists and managers still have to rely on statistical models for practical applications.

Although potentially skillful, the existing statistical forecasting methods entail several unresolved aspects. One such issue is that they lack a systematic principle for predictor screening. Even though experienced researchers are able to identify significant predictors in specific areas, predictor patterns are often too unique and complex to be readily discernible. Furthermore, statistical model forecasts are often expressed in forms that cannot be readily incorporated in impact assessment and decision making tools. This is the case of tercile type of probabilistic forecasts issued by various organizations (e.g., the CPC and IRI). Stakeholders and decision makers are generally unable to extract useful quantitative information from such climate forecasts to drive hydrologic and water resources models. This dichotomy between forecasting advances and water resources management needs is unfortunate, as many studies have demonstrated that considerable benefits can be accrued from skillful climate forecasting in several fields, including agriculture (Hansen et al., 2006; Jones et al., 2000; Mjelde et al., 1998), health care

(Thomson et al., 2006), socio-economics (Adams et al., 1995), and water resources management (Georgakakos et al., 1998, 2011a, b).

## 1.1 Objectives and Scope

This research aims to improve statistical climate forecasting and resolve its aforementioned weaknesses. More specifically, the research objectives and scope are as follows:

- To develop a novel method able to systematically identify the most relevant sea surface temperature (SST) predictors for the forecasting of hydro-climatic variables;
- To apply and demonstrated the value of the new method to the forecasting of hydro-climatic variables in various world regions and assess its skill;
- To understand the underlying physics of ocean-atmospheric interactions in the application regions; and
- To conceptualize an operational forecasting framework to support better water resources management.

## 1.2 Thesis Organization

The thesis has seven chapters organized as follows. Chapter 2 reviews the relevant literature on teleconnection patterns and climate forecasting. Chapter 3 summarizes all data used in this thesis and details the forecasting method. Chapters 4, 5, and 6 present case studies that apply the proposed method to the forecasting of seasonal precipitation over the southeastern United States and East Africa, and the forecasting of streamflow-

related variables in the Yangtze and Congo Rivers. Chapter 7 concludes this thesis with a summary of research contributions and major findings, and recommendations for additional future work.

# CHAPTER 2

# LITERATURE REVIEW

The development of a reliable hydro-climatic forecasting method requires the in-depth understanding of the dominant sources of climate variability as well as the contemporary trends in climate forecasting modeling. Section 2.1 reviews the relevant literature on climate variability with a specific aim of introducing various teleconnection patterns. Section 2.2 reviews many dynamical and statistical climate forecasting models developed in recent decades. Section 2.3 summarizes key messages from the literature review and addresses some research opportunities.

## 2.1 Climate Variability and Teleconnection Patterns

Atmospheric scientists have found that most of the dominant sources of climate variability are naturally recurring and persistent phenomena referred to as "teleconnection patterns." Generally, a teleconnection is a significant temporal correlation in a chosen variable between two far locations (van den Dool, 2007). Teleconnection patterns characterized by low-frequency variability can be associated with some large-scale oscillations in atmospheric dynamics (e.g., planetary waves, jet streams, and monsoons). This may subsequently influence temperature, rainfall, storm tracks, and other meteorological phenomena over vast areas. Therefore, scientists often rely on prominent teleconnection patterns to predict climatic conditions in remote locations with lead times from weeks to months (Chiew et al., 1998; Goddard et al., 2001; Palmer and Anderson, 1994).

In the late 19th to the early 20th centuries, meteorologist Sir Gilbert Walker began describing teleconnections to reference the correlation of sea-level pressure (SLP) between Tahiti and Darwin, which has become known as the Southern Oscillation (SO). In the late 20th century, with improved observations and identification techniques, researchers reinvestigated former teleconnection patterns throughout the troposphere (Wallace and Gutzler, 1981) with several updated extensions (Barnston and Livezey, 1987). In the following subsections, some of the most prominent teleconnection patterns as well as their effects on global climate are reviewed.

### *2.1.1 El Niño-Southern Oscillation*

The most prominent teleconnection pattern in the equatorial Pacific, which contributes the greatest amount of thermal energy to global climate variability, is El Niño. The discovery of El Niño, Spanish for "the boy," referring to the Christ child, dates back hundred years. The earliest record of El Niño was in the 19th century, when Peruvian fishermen noticed periodic warming behavior in the Pacific near South America around Christmas. Regardless of such warming, what initially drew meteorologists' attention was the adhering effect: the zonal anomaly of SLP in the tropical Pacific (i.e., the SO). Thus, the SO was extensively studied by meteorologists in the early 20th century (Walker, 1923; Walker and Bliss, 1932). Not until the 1960s did Jacob Bjerknes (Bjerknes, 1969) demonstrate that the SO was actually the atmospheric component of El Niño. Nowadays, these phenomena are jointly referred to as the El Niño-Southern Oscillation (ENSO).

Although ENSO does not have a single definition (Trenberth, 1997), for quantitative purposes, scientists have defined several regions in the equatorial Pacific to monitor variations of SST anomalies (SSTA) such as Niño 3, Niño 4, and Niño 3.4 [or the Oceanic Nino Index (ONI)] indices. In addition to the univariate measure of SSTA, scientists also derived the Multivariate ENSO Index (MEI) (Wolter and Timlin, 1993, 1998). Generally, when the running means of these indices are higher than a certain threshold (e.g., 0.5˚ for ONI), an El Niño event occurs. On the other hand, when the running means of these indices are lower than a certain threshold, the counterpart, La Niña, occurs.

The variation of ENSO can cause a variety of impacts. The primitive impact of ENSO on the Pacific fishing industry is due to anomalous warming SSTs that may suppress the upwelling of nutrient-rich cold water. In fact, ENSO has a more profound influence on the global-scale climate. While the Eastern Pacific undergoes anomalous warming (El Niño) or cooling (La Niña), the ordinary Walker circulation associated with wind patterns alters, thereby causing the typical tropical weather system to shift. In response, the climate in many tropical and extratropical regions experiences significant variations. These related global variations and composite patterns, especially articulated by precipitation and temperature, have been documented in numerous studies (e.g., Dai and Wigley, 2000; Harrison and Larkin, 1998; Kiladis and Diaz, 1989; Ropelewski and Halpert, 1987, 1989, 1996).

In contrast to the general understanding of the impacts of ENSO on the global scale, numerous studies have emphasized specific regions, revealing more detailed interactions and possible mechanisms. Such regions include Africa (Nicholson and Entekhabi, 1986; Nicholson and Kim, 1997), Australia (Power et al., 1999), Canada (Shabbar et al., 1997), East Asia (Wang et al., 2000), Europe (Fraedrich and Müller, 1992), India (Rasmusson and Carpenter, 1983), North America (Ropelewski and Halpert, 1986; McCabe and Dettinger, 1999), and South America (Rogers, 1988). Most of these regional studies intriguingly reached a consensus that opposite climatic conditions take place when the phase of ENSO changes. This opposite correspondence is clear evidence that tropical oscillations can shift climatic conditions in extratropical regions.

Apart from the general discussions focusing on the relationship between ENSO and global/regional climatic conditions such as temperature and precipitation patterns, manifold studies with different perspectives have been conducted. For example, some research attempted to connect ENSO to other important meteorological phenomena such as hurricane/tropical cyclone activity (Bove et al., 1998; Chan, 1985; Gray, 1984a) and monsoon variability (Lau and Nath, 2000; Webster and Yang, 1992). Other research of the ENSO influence emphasized extreme events such as extreme temperature, precipitation, and surface runoff, which are more likely to result in natural hazards. In comparison with shifts in their central tendencies, these events in high (or low) percentiles also shift accordingly with the phase changes in ENSO (Cayan et al., 1999; Gershunov and Barnett, 1998). Other examples of the various impacts of ENSO, including agricultural, health, water resources, and socio-economics in many world

regions have been documented in Diaz and Markgraf (2000), Glantz (2001), and Glantz et al. (1991).

As a result of the extensive impacts of ENSO, one of the consequential questions is its predictability. To simplify the problem, numerous studies have utilized the gross indices (e.g., Niño 3.4) as predictands or skill measures to optimize their prediction models. With years of development, either dynamical forecasting models (Cane et al., 1986; Chen et al., 1995; Ji et al., 1998) or statistical forecasting models (Barnston and Ropelewski, 1992; Xu and von Storch, 1990; see Subsection 2.2 for further discussions of two forecasting categories) outperformed persistence forecasting models, with the capability to forecast even with the lead-time of 12 months or more (Latif et al., 1994, 1998). Nevertheless, other scientists launched criticisms, claiming that the ENSO forecasts contained ambiguities and its skill had not yet been operationally validated as most models failed to forecast the very strong 1997-98 El Niño (Barnston et al., 1999; Landsea and Knaff, 2000). In addition, researchers have noticed a "springtime barrier" that may drastically deteriorate forecasting skill if lead-time is extended through the boreal spring (Webster and Yang, 1992). Thus, the improvement in ENSO forecasting still requires continuous effort.

Even though the warm anomaly in the eastern tropical Pacific has become the canonical indicator for El Niño, with more available data, especially from satellites, scientists have recently discovered a possible mutation pattern distinct from the conventional pattern. This new pattern is referred to as Central-Pacific El Niño, Dateline El Niño, or El Niño

Modoki (Ashok et al., 2007), in which a warming center is situated near the Date Line longitude rather than the typified eastern tropical Pacific. Because of the displacement of the warming center, analyses of seasonal associations carried out by a bootstrap technique have already shown that the Dateline El Niño would result in distinct impacts regionally and globally (Larkin and Harrison, 2005a, b). If numerical models are credible, some of their preliminary evidence has shown that the possible cause of El Niño Modoki is ongoing global warming (Yeh et al., 2009). Despite the inherent uncertainty of global warming, the Earth's climate experiences long-term modulation and variation. Therefore, the need to understand and predict these changes has become increasingly urgent.

### *2.1.2 North Atlantic Oscillation*

The most prominent and recurring extratropical teleconnection pattern in the Northern Hemisphere (NH) is the North Atlantic Oscillation (NAO). Although the earliest description of the NAO was noted by seafaring Scandinavians over 100 years ago, the initial meteorological documentation of the NAO was in the 1920s, approximately the time when the SO was discovered (Walker, 1924; Walker and Bliss, 1932). Unlike the SO, continuous and conspicuous research of the NAO had not been carried out until the late 1970s, when Rogers and other scientists published a series of papers pertaining to discussions about the "seesaw" between winter temperatures in Greenland and those in Northern Europe (Rogers and van Loon, 1979; van Loon and Rogers, 1978), which motivated many following endeavors (e.g., Lamb and Peppler, 1987). Similar to the indices of ENSO, the indices of the NAO (e.g., Rogers, 1984) are usually defined as the difference between the surface pressure of one location and that of another. For instance,

the widely-used Hurrell index is based on the difference between the normalized SLP in Lisbon, Portugal and that in Stykkisholmur/Reykjavik, Iceland (Hurrell, 1995).

Generally, the NAO refers to the meridional seesaw, or the dipole structure, typically associated with the SLP field of the north center near Iceland and the south center near the Azores across the Iberian Peninsula. The spatial structure of the NAO can be obtained through typical techniques such as simple one-point correlation maps (e.g., Wallace and Gutzler, 1981), leading empirical orthogonal functions (EOFs; e.g., Barnston and Livezey, 1987), and cluster analysis (e.g., Cassou et al., 2004), which are suitable for other teleconnection identifications. In spite of these slight differences, these techniques all signify two phase alterations in the NAO pattern. The "positive phase" of the NAO pattern indicates anomalously low sub-polar pressure (or geopotential height) and anomalously high subtropical pressure. The negative phase indicates the opposite pattern of either pressure or height over the same regions. Barnston and Livezey (1987) noticed this pattern had the largest amplitude (i.e., anomalous difference in SLP) during the boreal winter months, but a similar dipole structure may appear throughout the other months.

Due to the enhanced circulation of sub-polar low pressure during the positive NAO phase, stronger northerlies in the western side of this system may carry colder air that decreases land temperatures and SST over the northwest Atlantic. Meanwhile, stronger westerlies or south-westerlies at the east side of this system induce warm advection that increases land temperatures over Europe and further downstream. Along with the temperature

variations, the same system would further push the storm track northeastwardly so that enhanced storminess would be expected from Newfoundland into northern Europe (Hurrell and van Loon, 1997; Rogers, 1997; Serreze et al., 1997; Trigo et al., 2002). In contrast, during the negative NAO phase, both the opposite situation (e.g., northern Europe would face drier conditions) and the atmospheric blocking (Croci-Maspoli et al., 2007; Shabbar et al., 2001) would take place.

Because the background variability (i.e., changes in the mean flow and storminess) associated with oscillations of the NAO phase is prominent, studies have focused considerable attention on understanding and interpreting of how regional climate and water systems respond to the transport and convergence of atmospheric moisture and heat. For instance, such regional studies took place in Sweden (Chen and Hellström, 1999), the Middle East (Cullen et al., 2002), the Iberian Peninsula (Trigo et al., 2004), and the northeastern United States and Canada (Wettstein and Mearns, 2002). Detailed accounts of more NAO influences, including those affecting the SST, the subsurface ocean, sea ice, ecology, and the economy, have been documented in Hurrell et al. (2003) and a more recent paper (Hurrell and Deser, 2009).

### *2.1.3 Pacific/North American Pattern*

Another key teleconnection pattern in the NH is the Pacific/North American (PNA) pattern, which dominates many weather conditions in North America. The PNA is comprised of four teleconnection centers of 500-mb (Wallace and Gutzler, 1981) or 700-mb geopotential height fields (Barnston and Livezey, 1987). Similar to the NAO pattern,

these centers vary in pairs corresponding to the phase changes in the PNA pattern. The positive (negative) phase of the PNA pattern is defined as two above-average (below-average) height centers located in the vicinity of Hawaii (20˚N, 160˚W) and Alberta (55˚N, 115˚W), along with two below-average (above-average) height centers located in the vicinity of the northern Pacific Ocean (45˚N, 165˚W) and the Gulf Coast region (30˚N, 85˚W). The linear combination of the normalized height anomalies at these centers specifies the PNA index. Several other studies omitted the Hawaii center to exclude the barotropically unstable mode excited by anomalies in the East Asian jet stream (Leather et al., 1991; Leather et al., 1992; Yarnal and Diaz, 1986). The centers of the PNA pattern weave Rossby-like wavetrains that dominate the meridional and zonal anomalies of the jet stream. As a result, the disparate steering flow patterns have a significant effect on temperature and precipitation patterns over the United States, East Asia, and other regions in the Pacific basin (Leather et al., 1991).

### *2.1.4 Indian Ocean Dipole*

Similar investigations into identifying teleconnection patterns in the Indian Ocean had not been made until the very end of the 20$^{th}$ century. Saji et al. (1999) showed a dipole-like equatorial zonal mode later known as the Indian Ocean Dipole (IOD), which dominates large portions of climate variability in the Indian Ocean basin. The IOD was revealed because of the peculiar strong and persistent westward wind anomalies in the Indian Ocean during 1994 (Vinayachandran et al., 1999). By using composite analysis, researchers discovered that this peculiar wind pattern begins with the preceding cool anomalies of the SST in the southeastern tropical Indian Ocean, followed by a cold

intensification and gradual warm-up on the western side of ocean, which eventually results in the strongest gradient of the dipole structure and the wind pattern in October. This dipole, characterized by cooling (warming) in the eastern ocean and warming (cooling) in the western ocean, is referred to as the positive (negative) phase of the IOD. The phase changes in the IOD imply the anomaly of moisture transport, which impacts many regions significantly. For instance, the positive phase of the IOD would overwhelmingly enhance the short rains from October through November in East Africa (Behera et al., 2005); however, at the same time, the anomalous anticyclone on the eastern side may cause subsidence and induce droughts in Australia (Ashok et al., 2003).

### *2.1.5 Other Patterns Associated with Climate Variability*

Most teleconnection patterns reviewed in the previous subsections are found to be the first/leading modes according to the EOF analysis, which indicates dominant impacts on regional climate variability especially during the winter season. **Figure 2.1** illustrates some of the main teleconnection patterns. However, some other patterns such as the West Pacific Oscillation (WPO), the Tropical/Northern Hemisphere (TNH) pattern, and the East Atlantic (EA) pattern (Barnston and Livezey, 1987), which do not appear as leading modes, may still have substantial correlations with local temperature and precipitation. Beyond the winter season, because the atmospheric circulation is not favorable to the formation of large-scale patterns, relatively fewer teleconnection patterns can be identified. As a result, these patterns, appearing during the weakening periods [e.g., the North Pacific (NP) pattern (Barnston and Livezey, 1987; Bell and Janowiak, 1995)], have received only moderate attention.

If the temporal scale used in the pattern identification is extended to a decadal scale, other patterns referred to as "interdecadal modes" can be observed. For example, the Pacific Decadal Oscillation (PDO), a predominant source of Pacific climate variability, is characterized by a long-lived ENSO-like pattern that shifts phases with a period of at least 15 to 25 years (Mantua et al., 1997; Mantua and Hare, 2002; Zhang et al., 1997). Another example is the Atlantic Multidecadal Oscillation (AMO), signified with long-term SST variations with a period of 65 to 70 years (Schlesinger and Ramankutty, 1994). Both the PDO and the AMO have had widespread impacts on regional and global climate system such as hurricanes and ecosystems such as water resources and marine fishery resources (Enfield et al., 2001; Goldenberg et al., 2001; Mantua et al., 1997; Mantua and Hare, 2002; McCabe et al., 2004)

Even though many interannual and interdecadal modes have been identified, the underlying mechanisms remain obscure. This problem causes mixed characteristics and effects among patterns, which are then less effective at predicting climatic conditions. Because of such difficulties, understanding climate variability through exploring new patterns and determining their underlying mechanisms is still a challenge.

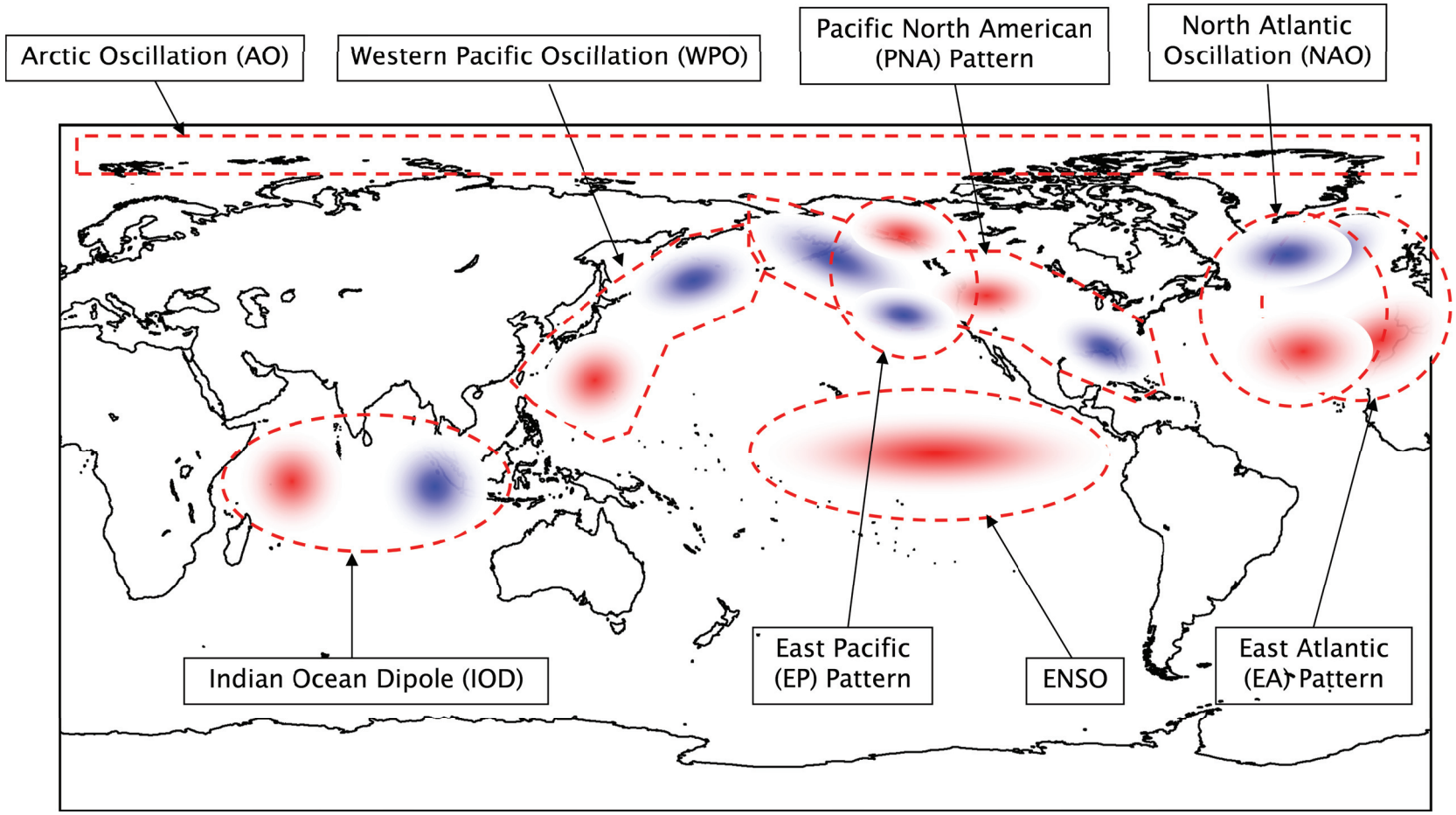



**Figure 2.1:** Illustration of prominent teleconnection patterns.

## 2.2 Climate Forecasting

Even though various forecasting methods have been invented in the last several decades, climate forecasting can generally be classified into two broad categories. Since high-performance computing became available, the first, which solves discretized primitive equations numerically, is usually referred to as "dynamical forecasting." In contrast, the second employs empirical or statistical techniques that directly correlate the local climate with large-scale variables from general circulation diagnostics. This type of climate forecasting is usually referred to as "statistical forecasting." In this context, these two forecasting categories are reviewed next.

### *2.2.1 Dynamical Forecasting*

Before the weather system is overwhelmed by chaotic effects (Lorenz, 1963), short-term weather is predictable using dynamical models that integrate primitive equations over time with proper initial conditions and boundary conditions. Even though the capability of precisely forecasting mid- to long-term weather/climate is impeded by the chaotic evolution of initial conditions, seasonal-mean conditions could be depicted by dynamical forecasting forced with ideal (observed) lower-boundary conditions (i.e., SST) typified by noticeable persistency. This modeling framework is often referred to as a "two-tiered" approach (Bengtsson et al., 1993). For example, Graham et al. (2000) assessed seasonal predictability using several atmospheric general circulation models (AGCMs) involved in the European PRediction of climate Variations On Seasonal to interannual Time scales (PROVOST) project. They integrated AGCMs with prescribed observed SSTs and evaluated their probabilistic skill for the three-month mean 850 hPa temperature and the

three-month mean precipitation accumulation in regions including the tropics, the northern extratropics, Europe, and North America. Whereas Graham et al. emphasized the assessment of skill, Palmer et al. (2000) performed a decision-model analysis of seasonal ensemble integrations from the PROVOST project. Both papers drew identical conclusions: the skill and the value of forecasts were enhanced during the ENSO years; the multi-model ensembles were substantially more skillful. Similar to the PROVOST project, the Dynamical Seasonal Prediction (DSP) project investigated five state-of-the-art U.S. AGCMs forced with a predicted SST (Shukla et al., 2000). Although the emphasis of DSP is on the PNA region, results also show that its significant predictability is attributed to the larger anomaly of the tropical SST (i.e., ENSO).

Another example of utilizing a two-tiered approach was demonstrated by the IRI seasonal climate forecast system. Mason et al. (1999) summarized how the IRI used a two-tiered-based multimodel system to begin issuing seasonal forecasts in 1997. The forecasting product of the IRI is featured in probabilistic three-category forecasts for total precipitation and mean temperature covering most continental regions. Afterwards, Goddard et al. (2003) evaluated the IRI forecasts from 1997 to 2001. Because of the strong ENSO events during this period, they found that seasons and regions with presumable predictability had enhanced forecasting skills. More recently, Barnston et al. (2010) verified IRI seasonal climate forecasting for a longer period from 1997 to 2008. The 11-year forecasts generally presented consistent predictive skill compared to previous assessments; with improvements such as an increase in the number of

constituent AGCMs (from three to seven) and a refinement of multiple SST scenarios, the IRI forecasting system has become more reliable.

Since two-tiered approaches highly depend on the SST boundary condition, considerable work has been done to investigate the influence of different SST configurations on AGCMs. This problem was one of the most significant motivations for the Atmospheric Model Intercomparison Project (AMIP; Gates, 1992). More specifically, an AMIP-type simulation examines two measures of AGCM performance regarding SST forcing: skill and reproducibility (Moron et al., 1998). While skill represents the capability of the model to resemble the observed climate with prescribed SST forcing, reproducibility represents the capability of the model to converge on similar patterns with varied initial atmospheric conditions. The significance of both measures in AGCM simulations is fundamental to conducting realistic forecasts forced with persisted or predicted SST anomalies (Goddard and Mason, 2002; Li et al., 2008). However, because of inaccurate predictions of SST and incomplete physics in AGCMs, the performance of two-tiered approaches is still limited to certain regions and seasons.

Scientists attributed the major weakness of two-tiered approaches to the lack of dynamical interaction between the atmosphere and the ocean. The possibility of the mutual evolving of boundary conditions can be fulfilled by coupling ocean-atmosphere general circulation models (CGCMs). Because the process of configuring the prescribed SST is no longer required, the CGCM is also referred to as a "one-tiered" approach. Development and assessment of CGCMs have been taken place only in a few operational

centers such as the NCEP (Ji et al., 1998; Kanamitsu et al., 2002; Saha, et al., 2006) and the ECMWF (Stockdale et al., 1998; van Oldenborgh et al., 2005) because of the enormous resource requirements. However, many CGCM experiments were still conducted in several regions. For example, based on the results from the PROVOST project, the Development of a European Multimodel Ensemble System for Seasonal to Interannual Prediction (DEMETER) project was conceived to examine seven CGCMs used in European countries (Palmer et al., 2004). Even though the one-tiered framework seems to be more physically rational, CGCMs have to be employed with caveats such as the "drift" SST problem accompanied with the integration process. Because of such caveats, few operational centers (e.g., IRI) are very prudent to progress through a totally one-tiered system.

Regardless of the choice of either a two- or a one-tiered approach, dynamical forecasts have intrinsic problems, including systematic biases and limited skill. The systematic biases can usually be ameliorated by ensemble forecasting (Anderson, 1996; Barnett, 1995; Branković and Palmer,1997; Sivillo et al., 1997; Palmer and Anderson, 1994) or multimodel ensemble forecasts (Krishnamurti et al., 2000; Rajagopalan et al., 2002; Robertson et al., 2004). Furthermore, numerous statistical correction methods have been implemented to compensate for limited skill (Feddersen et al., 1999; Mo and Straus, 2002; Tippett et al., 2003, 2005). Nonetheless, despite the understanding of systematic biases and the adjustment of skills, dynamical models still cannot produce satisfying, skillful forecasts. In fact, the problem of skill deficiency can be further exemplified by a recent study done by Lavers et al. (2009), who assessed the predictability of the aforementioned

NCEP Climate Forecast System (Saha et al., 2006) and seven models from the DEMETER project. Their results showed that if the lead time increases to 31 days, predictions of temperature have noticeable skill primarily located over the tropics, but predictions of precipitation have insignificant skill. In addition, with the same lead time, predictions of both temperature and precipitation have negligible skill over land masses. Thus, to provide useful forecasts, dynamical models need significant improvement.

### *2.2.2 Statistical Forecasting*

The second category of climate forecasting, which estimates local climatic condition (predictand) based on predictor candidates either from general circulation diagnostics or purely statistical techniques, is statistical forecasting (Hastenrath, 1995). In most cases, statistical forecasting begins with the common process of predictor screening. Similar to the two-tiered dynamical approaches, predictors are heavily dependent upon the SST boundary field, which has been proven to induce seasonal fluctuations in atmospheric circulation even in the mid-latitude (Frankignoul, 1985). In contrast to dynamical forecasting, which utilizes the SST boundary forcing in numerical models to solve exact temporal and spatial transitions of weather/climate regimes, statistical forecasting projects variations in the SST field onto the predictand field through a statistical predictor-predictand (causal) relationship with tuned parameter(s); furthermore, the estimation of predictand is usually subject to some probabilistic distribution.

As a result, many investigators adopt the SST as a sole or dominant predictor to develop statistical causal relationships in different world regions. For instance, Folland et al.

(1991) used multiple linear regression (MLR) and linear discriminant analysis (LDA) to predict Sahel rainfall based on selected SST EOFs as predictors. Mutai et al. (1998) chose three rotated SST EOFs as predictors and applied MLR and LDA to predict the East Africa short rains (during October to December). Uvo et al. (1998) employed singular value decomposition (SVD) to examine the concurrent correlations between the monthly tropical Pacific and Atlantic SSTs and precipitation in the northeast region of Brazil and then assessed the possibility of seasonal forecasts. Colman and Davey (1999) calculated EOF patterns of the north Atlantic SSTA used in regression methods to predict summer temperatures, SLP, and rainfall in Central England and Europe. Drosdowsky and Chambers (2000) used rotated principal component analysis to represent the near global SSTA patterns align with LDA to forecast seasonal rainfall in Australia. Sharma (2000) and Sharma et al. (2000) presented one specific criterion as a basis for identifying predictors from gross indices including NINO3, NINO3.4, and SOI together with SSTs for seasonal to interannual rainfall forecasts. Markowski and North (2003) used MLR to correlate the calculated EOFs of prescribed Pacific and the Gulf of Mexico SST regions with the precipitation over the United States. Diro et al. (2011a, 2011b) examined the linkage between global SSTs and the Kiremt rainfall (during June to September) over Ethiopia by using composite analyses based on ERA-40 reanalysis data and then develop a seasonal forecasting system based on LDA and MLR.

In some cases, including predictors in addition to SSTs may be beneficial, especially when the upper limit of predictability may result from the sole dependency on the SST field (Westra and Sharma, 2010). For Instance, Philippon et al. (2002) and later

Hastenrath et al. (2004) introduced many "circulation indices" such as wind, pressure, omega vertical motion, divergent zonal flow, specific humidity, in addition to SSTs as predictors of East Africa short rains. Grantz et al. (2005) selected the snow water equivalent, 500-mb geopotential height, and SSTs to develop the local regression and ensemble forecasting for the prediction of streamflow in the Truckee-Carson River flow. Rajeevan et al. (2007) developed statistical models that selected predictors from oceanic and atmospheric fields including SSTA, surface pressure anomaly, surface air temperature anomaly, warm water volume, and zonal wind anomaly at 850 hPa in specific regions to forecast the southwest monsoon rainfall over India. However, even though these studies took advantage on additional predictors to enhance forecasting skills, some scientists (e.g., Goddard et al., 2001) argued that predictability from atmospheric predictors is still the derivative of SSTs.

Among the statistical forecasting models, regression models are probably the most common in climate applications (Goddard et. al., 2001; Zwiers and von Storch, 2004). As a possessive category, canonical correlation analysis (CCA), which is one of the operational methods adopted by the CPC, expands simple regression, one variable upon another, to multidimensional vectors (Barnett and Preisendorfer, 1987; Barnston and Ropelewski, 1992; Barnston, 1994). CCA is guaranteed to explain the maximum variance in the predictand field through simply finding the optimum linear combination of the predictor fields, so it has been extensively adopted in many climate forecasting studies (Barnston and He, 1996; Yu et al., 1997; Johansson et al., 1998; Ward, 1998; Landman and Mason, 1999; Wilks, 2008). Despite the strength of CCA, including too

many variables as predictors may result in forecast bias. It is believed that one of the prime reasons for using CCA is that it provides side-by-side patterns that pinpoint where forecasting skills occur. For this reason, van den Dool (2007) stated CCA is more like a tool of finding coupled modes. Nevertheless, patterns and major causal relationships derived from CCA are not always easily interpretable or isolated since too much information may be involved.

Instead of using linear modeling, some recent studies advocate nonlinear modeling since it offers more flexibility and possibly better performance; more importantly, nonlinearity may be closer to the nature of climate mechanisms. Among these nonlinear models, neural network (NN) may be the most popular one (Hsieh and Tang, 1998; Tangang et al., 1998; Tang et al., 2000). In addition, other models, including the nonlinear generalizations of CCA (Hsieh, 2001) or discriminant analysis (Mason, 1998) and specific techniques from machine learning (Lima et al., 2009; Lima and Lall, 2010), were also employed in several studies. The reviewed climate forecasting studies that utilized statistical techniques are summarized in **Table 2.1**.

**Table 2.1:** List of statistical climate forecasting studies reviewed in the thesis (authors are in alphabetical order).

| Author(s) | Predictor Description* | Predictand Description* | Forecasting Method* |
|---|---|---|---|
| Barnett and Preisendorfer (1987) | Temperature persistence, SST, and SLP | Surface air temperatures in the United States | CCA |
| Barnston and Ropelewski (1992) | Seasonal mean SLP (near global), SST in the ENSO region | SST in several regions of the tropical Pacific (ENSO) and Indian oceans | CCA |
| Barnston (1994) | SST (near global), Northern Hemisphere 700-mb height, and prior values of predictand field | Surface temperature and precipitation and 700-mb geopotential height in the Nothern Hemisphere | CCA |
| Barnston and He (1996) | SST (near global), Northern Hemisphere 700-mb height, and prior values of predictand field | Surface temperature and precipitation in Hawaii and Alaska | CCA |
| Bierkens and van Beek (2009) | The NAO index | Winter and summer discharge of European rivers | A hydrological model forced with resampling ERA-40 data based on the NAO index |
| Chan et al. (1998, 2001) | ENSO and ENSO-related indecis, parameters related to the large-scale circulation, and trend of the interannual variations in TC activity | Seasonal tropical cyclone activity over the western North Pacific and the South China Sea | Projection-pursuit regression |
| Chu et al. (2007) | SSTs, SLPs, vertical wind shear, relative vorticity, and precipitable water in specific times and regions | Seasonal tropical cyclone activity over the central North Pacific | Bayesian regression |
| Colman and Davey (1999) | SST EOFs (the North Atlantic) | Summer temperatures, SLP, and rainfall in Central England and Europe | MLR |
| Diro et al. (2010a, 2010b) | SST regions identified by correlation analysis | Kiremt rains in Ethiopia | MLR and LDA |
| Drosdowsky and Chambers (2000) | PCA SSTA patterns (near global) | Seasonal rainfall in Australia | LDA |

**Table 2.1:** *Continued.*

| Author(s) | Predictor Description* | Predictand Description* | Forecasting Method* |
|---|---|---|---|
| Fan and Wang (2009) | Sea ice coverage, 1000-hPa temperature, 850-hPa vorticity, the magnitude of the wind shear between 200 and 850 hPa, and SLP in specific times and regions | Typhoon frequency over the western North Pacific | MLR |
| Folland et al. (1991) | Selected SST EOFs (global) | Seasonal rainfall in Sahel | MLR and LDA |
| Grantz et al. (2005) | Snow water equivalent, 500-mb geopotential height, and SST indices | Streamflow in the Truckee-Carson River | Local regression |
| Gray (1984b) | Quasi-Biennial Oscillation (QBO) direction and change, the El Niño index, and the April-May SLPA correction factor | Atlantic seasonal hurricane frequency | MLR |
| Gray et al. (1992) | QBO zonal winds, August-September western Sahel and August-November Gulf of Guinea rainfall | Atlantic seasonal hurricane frequency | Least-absolute deviation regression |
| Gray et al. (1994) | QBO zonal winds and vertical shear; West African rainfall, SLP, and temperature; Caribbean 200-mb zonal winds and SLP; ENSO SSTs and the SOI | Atlantic seasonal tropical cyclone activity | Least-absolute deviation regression |
| Hamlet and Lettenmaier (1999) | ENSO and PDO signals | Columbia river streamflow | Resampling data according to six predefined PDO/ENSO categories |
| Hastenrath et al. (2004) | Indices of surface westerlies, divergent zonal flow, 500-hPa omega vertical motion, pressure and SST difference, and SOI | Short rains in East Africa | MLR |
| Hsieh (2001) | The tropical Pacific SLP | The tropical Pacific SST | Nonlinear CCA using NN |
| Johansson et al. (1998) | The predictand field itself, local SST in the northern European area, Northern Hemisphere 700-hPa height, and quasi-global SST | Surface temperature in northern Europe | CCA |

**Table 2.1:** *Continued.*

| Author(s) | Predictor Description* | Predictand Description* | Forecasting Method* |
|---|---|---|---|
| Fan and Wang (2009) | Sea ice coverage, 1000-hPa temperature, 850-hPa vorticity, the magnitude of the wind shear between 200 and 850 hPa, and SLP in specific times and regions | Typhoon frequency over the western North Pacific | MLR |
| Kim et al. (2010) | SST, outgoing long-wave radiation, and 850-hPa relative vorticity in specific times and regions | Summertime tropical cyclone activity over the East China Sea | Least-absolute deviation regression and Poisson regression |
| Klotzbach and Gray (2004) | Sep and Nov 500-mb geopotential height, ON SLP, Jul 50-mb equatorial zonal wind, Sep-Nov SLP, and Nov SLP in specific regions | Atlantic seasonal tropical cyclone activity | MLR and Poisson regression |
| Klotzbach (2008) | ON SST, Nov 500-mb geopotential height, Nov SLP in specific regions | Atlantic seasonal tropical cyclone activity | MLR |
| Kung and Sharif (1980) | 100- and 700-mb geopotential height, kinetic energy, zonal and meridional wind, and temperature at the forecast point | The onset of the Indian summer monsoon | MLR |
| Kwon et al. (2007) | SST, SLP, 850-hPa temperature, meridional wind, and 500-hPa geopotential height regions that show high correlations | Seasonal tropical cyclone activity over the western North Pacific | MLR |
| Landman and Mason (1999) | SSTs (global) | Summer rainfall in South Africa | MLR and CCA |
| Lima et al. (2009) | Canonical modes of the 20°C isotherm and lagged values of ENSO indices | Several ENSO indices | Maximum variance unfolding and PCA-based models |
| Lima and Lall (2010) | The NINO3 index and the canonical modes of the 20°C isotherm | Brazilian hydropower system | Maximum variance unfolding and PCA-based models |
| Lo et al. (2007) | The SOI and tropical Pacific SSTs | The onset of the north Australian wet-season | Linear logistic regression |

**Table 2.1:** *Continued.*

| Author(s) | Predictor Description* | Predictand Description* | Forecasting Method* |
|---|---|---|---|
| Lu et al. (2010) | 850-hPa relative vorticity, vertical wind shear, and SLP over the western and central North Pacific from the antecedent May | Tropical cyclone activity near Taiwan | Bayesian regression |
| Markowski and North (2003) | SST EOFs (prescribed Pacific and the Gulf of Mexico regions) | Seasonal precipitation in the United States | MLR |
| Mason (1998) | Principal components of SSTs in the Indian, South Atlantic and Pacific Oceans | South African rainfall | Nonlinear DA |
| Mutai et al. (1998) | Three rotated SST EOFs (global) | Short rains in East Africa | MLR and LDA |
| Nicholls (1979) | Darwin winter pressure | Tropical cyclone frequency | Correlation analysis |
| Nicholls (1984) | Darwin pressures | The onset of the north Australian wet-season | Linear logistic regression |
| Philippon et al. (2002) | Principal component, monsoon dynamics index, and 200-hPa meridional wind near South Africa | Short rains in East Africa | MLR and LDA |
| Sharma (2000) and Sharma et al. (2000) | NINO3, NINO3.4, and SOI together with SSTs identified by the partial mutual information criterion | Quarterly rainfall at Warragamba dam near Sydney Australia | Generalised Additive (regression )Model |
| Tangang et al. (1998) | Extended EOFs of the SLP field | SSTA in ENSO regions | NN |
| Tang et al. (2000) | Extended EOFs of the SLP and SST field | SSTs in ENSO regions | CCA, LR, and NN |

**Table 2.1:** *Continued*.

| Author(s) | Predictor Description* | Predictand Description* | Forecasting Method* |
|---|---|---|---|
| Uvo et al. (1998) | SSTs in the tropical Pacific and Atlantic Oceans | Monthly precipitation in Northeast Brazil | Lagged SVD |
| Ward (1998) | SSTs (global) | Summer rainfall in tropical North Africa | MLR and CCA |
| Wilks (2008) | Long (since 1880) and short (since 1950) Pacific SST data | North American temperatures | CCA and maximum covariance analysis |
| Yu et al. (1997) | SSTs in the Pacific Ocean | Rainfall in 10 stations at the U.S. Affiliated Pacific Islands | CCA and multivariate principal components regression |

*: Please refer to the text for the abbreviations.

### *2.2.3 Other Climate Forecasting Applications*

One of the climate forecasting applications is linked to streamflow forecasting. Some studies used dynamical approaches and then translated associated precipitation forecasts into streamflow forecasts through hydrologic models. For example, Wood et al. (2002) prepared bias-corrected and downscaled GCM forecasts for hydrologic modeling to generate streamflow forecasts for the eastern United States. Few years later, Wood and Lettenmaier (2006) applied the similar concept to the western United States. Other studies employed statistical approaches to bypass intervening processes and then correlated large-scale phenomena with hydrologic variables. For example, Hamlet and Lettenmaier (1999) used resampling technique to generate Columbia River streamflow forecasting according to predefined PDO/ENSO categories. Bierkens and van Beek (2009) devised a NAO-based seasonal prediction framework for European discharge.

Another notable application is seasonal forecasting of tropical cyclone (TC) or hurricane activity. Nicholls (1979) pioneered this application to examine tropical cyclone frequency over the Australian region. Another series of efforts aiming at Atlantic hurricane forecasting were made by the Tropical Meteorology Project at Colorado State University (Gray, 1984b; Gray et al., 1992, 1994; Klotzbach and Gray, 2004; Klotzbach, 2008). Forecasting models were also developed in other TC-affected regions such as western North Pacific (Chan et al., 1998, 2001; Kwon et al., 2007; Fan and Wang, 2009) and East Asia (Chu et al., 2007; Kim et al., 2010; Lu et al., 2010).

Another application especially important for agricultural development is to predict the onset and/or cessation of a rainy season. Numerous studies have attempted to address such task by using regression models in monsoon-prevailing regions, such as India and Australia (Kung and Sharif, 1980; Nicholls, 1984; Lo et al., 2007).

More climate forecasting applications related to agriculture are the predictions of crop yield. For instance, Cane et al. (1994) found that maize yield in Zimbabwe could be forecasted by the leading NINO3 index due to the significant correlation. According to Mavromatis et al. (2002), knowing ENSO conditions may have positive effects on "enhancing peanut yield (1 to 8%) and reducing yield variability (2 to 12%)" in northern Florida and southern Georgia. The DEMETER project (Palmer et al., 2004) also demonstrated how the seasonal climate prediction can be applied to predict crop yield. However, even though many assessments have acknowledged that seasonal climate forecasting is generally helpful to agriculture, continuous work is still necessary to fill gaps in these assessments (Meza et al., 2008).

### 2.3 Literature Review Summary and Research Opportunities

This chapter reviewed substantial studies related to the thesis work. In the first part, some of the most prominent teleconnection patterns, including ENSO, the NAO, the PNA, the IOD, other secondary patterns, and interdecadal modes, were reviewed. These recurring large-scale patterns are proven to be associated with climate variability in many world regions. Nevertheless, because progressive teleconnection patterns have high

temporal and spatial variations, oftentimes they can only provide approximate onward trends of regional climatic conditions.

The second part of this chapter reviewed many articles related to dynamical and statistical climate forecasting. Because dynamical models rely on the numerical computation of primitive equations, they are able to predict regional/global climatic conditions with the most plausible physics. In addition, with the advance of computing technology, dynamical models have evolved from a two-tiered (AGCMs), low-resolution approach into a one-tiered (CGCMs), high-resolution approach. Despite such solid physics and sophisticated computations, dynamical models are not capable of providing accurate forecasts for most extratropical terrestrial areas yet. On the other hand, a variety of statistical techniques (e.g., regression, SVD, CCA, and NN) that establish mathematical relationships between large-scale predictors and regional climatic predictands exhibits more satisfactory forecasting skills. Furthermore, the flexible choice of predictands can further manifest the practicality of statistical models.

Plentiful studies have shown applications and associated benefit that result from skillful forecasting practices. While statistical models can possibly generate such forecasts, they often cannot provide physical explanation. Therefore, incorporating the characteristics of prominent teleconnection patterns in developing a statistical model should be a research opportunity. However, a new approach should be capable of identifying more sensitive signals apart from the well-known patterns as reviewed so that it can better explain regional hydro-climatic variables.

Furthermore, since most of the existing forecasting products are presented in a contained format (e.g., tercile type of probabilistic forecasts), integrated work that derives beneficial quantity from raw forecasts becomes intangible. Thus, an imperative need is to design a reliable, convertible, and flexible forecasting product with demonstrated operational utility.

# CHAPTER 3

# DATA AND METHODOLOGY

This chapter first introduces various datasets used in the thesis and then interprets the new climate forecasting method. Section 3.1 summarizes an SST product, two precipitation datasets, other streamflow-related data, and a climate reanalysis dataset. Section 3.2 describes the predictor screening algorithm, comprising a specific predictor structure, multiple skill measures for predictors, and a phased predictor searching process. Section 3.3 describes the forecasting process, comprising linear regression and a cross-validation scheme.

## 3.1 Data Summary

### *3.1.1 Sea Surface Temperature*

Among several qualified SST products, this study adopts the Kaplan Extended SST V2 product, provided by the NOAA/OAR/ESRL PSD (Boulder, Colorado) on their website (http://www.esrl.noaa.gov/psd/), as the predictor field. Kaplan et al. (1998) produced monthly global SST by applying several statistical interpolation procedures to the United Kingdom Meteorological Office historical sea surface temperature (MOHSST5). The Kaplan SST has the advantage of a long data record from 1856 to the present. Satellite data available after 1981 were incorporated through the EOFs used in the Kaplan SST. The EOFs included the projection of the NCEP optimal interpolation (OI) analysis (Reynold and Smith, 1994) enabling the Kaplan SST to enhance its data quality without losing temporal homogeneity (pre- and post-1981). Although the resolution of the

Kaplan SST ($5^o$ x $5^o$) may not be as fine as other contemporary products, it is sufficient for this work that is interested in prominent teleconnection patterns. Furthermore, since some oceanic areas have sparse observations, very fine resolution data obtained from over-interpolation may actually increase uncertainty.

### *3.1.2 Precipitation*

One of the predictands in this thesis is regional precipitation. Similarly, the important requirements of regional precipitation for climatic research are sufficient data size and quality. Two such data sets are described next.

#### *3.1.2.1 CMAP Data*

One of the most widely-used, high-quality global precipitation data is the CPC Merged Analysis of Precipitation (CMAP) data. CMAP produces monthly (also available in pentad) global precipitation on a $2.5 \times 2.5$ degree grid based on an algorithm that merges observations from rain gauges with five kinds of satellite estimates (standard version) and with numerical model predictions (enhanced version). Because the merging algorithm was first described by Xie and Arkin (1997), the CMAP data is sometimes referred to as the Xie-Arkin dataset. The derived CMAP data is from 1979 to near present and available on the CPC FTP server (ftp.cpc.ncep.noaa.gov/precip/cmap/monthly/). In this study, the CMAP data is used to generate seasonal precipitation in world regions where no better precipitation product can be acquired.

*3.1.2.2 PRISM Products for the United States*

Precipitation gauge networks in the United States have relatively high density, thus providing spatial datasets of high quality. By using the PRISM (Parameter-elevation Regressions on Independent Slopes Model) climate mapping system, the PRISM Climate Group at Oregon State University produces 2.5-arcmin (4 km) grid-based climatic data, including precipitation, average maximum temperature, average minimum temperature, and average dewpoint for the entire United States. In addition to point data and a digital elevation model, the PRISM Climate Group further introduced a knowledge-based approach that accounts for rain shadows, coastal effects, and temperature inversions to produce their data (Daly et al., 2002). The monthly data is available on their website (http://www.prism.oregonstate.edu/) with a record from 1895 to the present. Because the PRISM Climate Group provides very fine precipitation data, spatial average precipitation in the region of interest (e.g., watershed) is sometimes used.

*3.1.3 Other Predictand Data*

In addition to regional precipitation, other hydro-climatic variables can also be used as predictands. Specifically, the inflow data at the Three Gorges Dam at the Yangtze River and the water level data at the Kinshasa station at the Congo River, both obtained from local authorities, are also used in the study.

*3.1.4 NCEP/NCAR Reanalysis Dataset*

Climate diagnostics that examine the patterns of atmospheric variables (e.g., pressure, geopotential height, and wind) in different vertical levels to elucidate plausible

mechanisms for regional climate processes are also employed in this work. Because climate diagnostics require a comprehensive and consistent data record for all atmospheric variables, reanalysis datasets created by assimilating climate observations (e.g., from ships, planes, buoys, stations, and satellites) into climate model simulations are usually used. In this study, the NCEP/NCAR Reanalysis dataset (Kalnay et al., 1996), available on the NOAA PSD website (http://www.esrl.noaa.gov/psd/), is adopted. This state-of-art dataset is based on the NCEP/NCAR Reanalysis 1 project, which generates 4-times daily, daily, and monthly values of various atmospheric variables in different vertical levels (17 pressure levels and 28 sigma levels) with a global coverage (on a resolution of $2.5 \times 2.5$ degree) from 1948/01/01 to the present.

### 3.2 Predictor Screening Algorithm

Generally, the proposed forecasting method consists of five key processes depicted in **Figure 3.1**. The first process is to define spatial and temporal domains of SST and hydro-climatic data introduced in the last section as well as to input necessary parameters for the algorithm. The second process is to identify significant SST predictors in the specific form of "dipoles" based on a screening algorithm interpreted in this section. The third process is to employ the identified predictors to generate ensemble-like forecasts. This process is composed of linear regression and a cross-validation scheme, which is described in Section 3.3. After forecasts are generated, the fourth process is to illustrate identified dipoles and to assess forecasting skills by using various measures introduced in Section 3.2.2. The fifth process is to perform climate diagnostics by using the reanalysis

data to generate composite maps of atmospheric variables in different levels. The last two processes are exemplified by case studies carried out in the subsequent chapters.

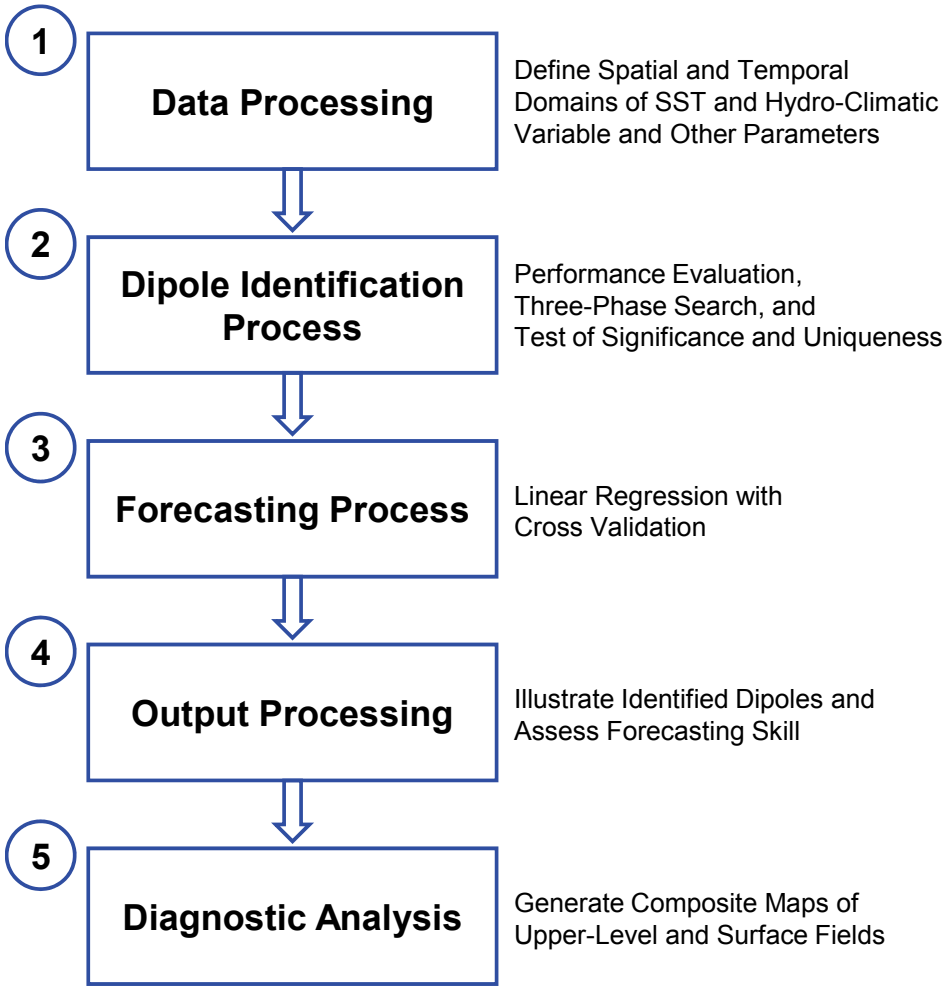



**Figure 3.1:** General flow chart of the proposed forecasting method.

*3.2.1 Dipole as a Predictor Structure: Motivation and Definition*

For the purpose of developing predictive relationships between teleconnection patterns and regional climate, scientists have defined and formulated indices for these phenomena by delineating fixed areas as sensitive regions (van den Dool, 2007). Typical examples include the Niño 3.4 Index (Trenberth, 1997), the NAO (North Atlantic Oscillation) Index (Hurrell, 1995), and the Dipole Mode Index (DMI) (Saji et al., 1999). Depending on the number of geographic centers, the fixed areas of these teleconnection indices are usually referred to as a monopole, a dipole, or a multipole. With more poles, a teleconnection mode may offer more flexibility that can better explain climate variability in certain regions. However, this kind of pattern may display too much physical complexity to be understood. Based on classic techniques such as EOF analysis, scientists find that most significant teleconnection modes are usually monopole-like or dipole-like systems (see **Figure 2.1**). Thus, a predictor screening algorithm that employs the dipole structure is proposed.

Dipole time series $D(t)$ is defined as the mathematical difference or addition of the attributes (i.e., SST) over two pixels or two clusters of pixels:

$$D_{\phi_1,\phi_2}(t) = Avg[\phi_1] \pm Avg[\phi_2], \quad (1)$$

where $\phi$ denotes the attribute of each pole, and *Avg* denotes the spatial average over the pole region. Each pole of a dipole should comply with two logical constraints: 1) If either of the poles contains multiple pixels, these pixels should be always contiguous and clustered; and 2) The poles of a dipole can be contiguous, but they cannot overlap or share any common pixels. Following the definition and constraints, possible dipole

configurations can be illustrated: **Figures 3.2(a)** and **3.2(b)** show that each pole of a dipole can contain a single pixel or a cluster of pixels; **Figure 3.2(c)** show that a tripole can form if the identified dipoles share a common pole; **Figure 3.2(d)** show that a dipole can degenerate into a monopole if each pole has the same sign and is contiguous. Thus, a dipole pattern is capable of generating various teleconnection patterns.

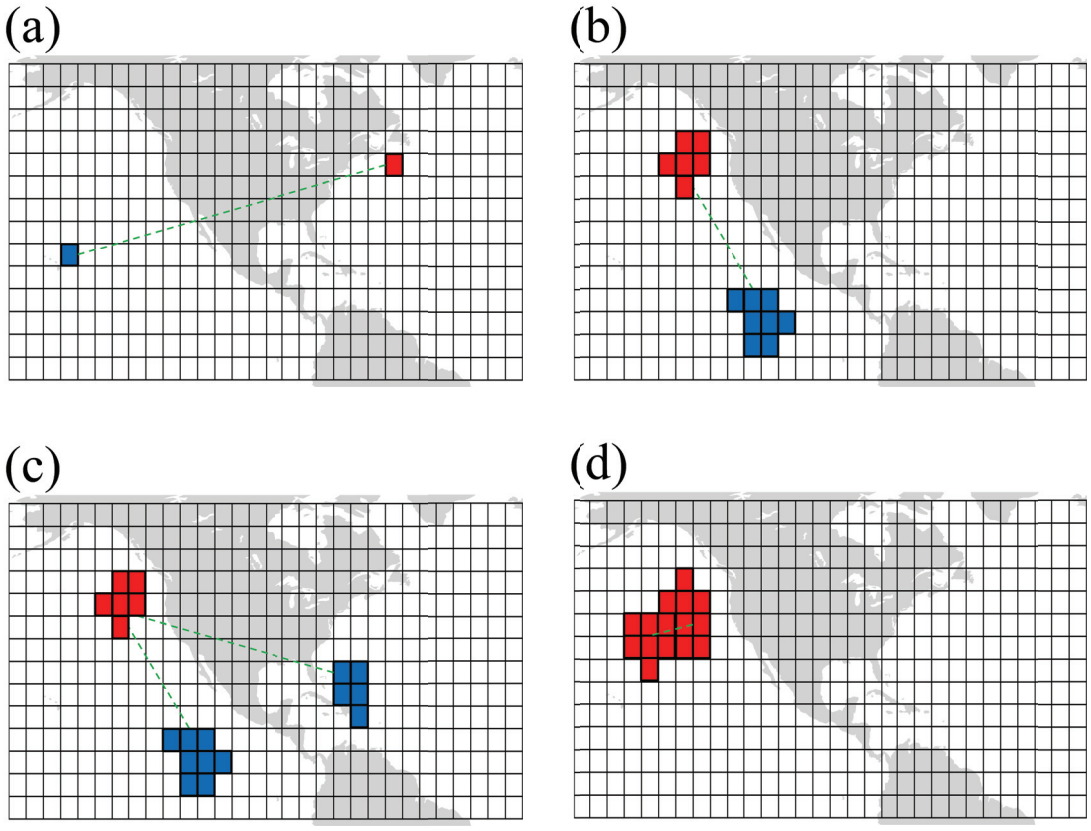



**Figure 3.2:** Illustration of different dipole configurations: (a) a single pixel in each pole; (b) a cluster of pixels in each pole; (c) a tripole (one common positive pole and two separated negative poles); (d) a monopole (contiguous poles with the same sign).

The development of the dipole teleconnection algorithm must address the following questions: 1) how to measure the predictive performance of dipoles? 2) how to determine the most adequate dipole sizes and geographic locations that produce the highest predictive performance? These questions are answered in the following subsections.

### *3.2.2 Performance Evaluation*

In a prediction context, the predictor is $D_{\phi_1,\phi_2}(t)$, while the predictand may be any external hydro-climatic variable, such as precipitation or streamflow, lake levels, and groundwater levels.

#### *3.2.2.1 Skill Score*

The conventional way to assess forecast skill is to begin with the development of an event contingency table, which summarizes the pairwise categorized information between the predictor and predictand. In addition, the size of a contingency table ($J$) can be as large as the data size permits. After a contingency table is developed, various types of skill scores recommended can be calculated (Wilks, 2006). This study adopts the Gerrity Skill Score (GSS) which possesses several advantages. The GSS (Gerrity, 1992), an "equitable skill score" developed by Gandin and Murphy (1992), has the following general form:

$$SS = \sum_{i=1}^{J}\sum_{j=1}^{J} p_{ij} s_{ij}, \tag{2}$$

where $p_{ij}$ is the element of performance matrix **P** and $s_{ij}$ is the element of scoring matrix **S** as shown in **Figure 3.3**. The entries $p_{ij}$ of matrix **P** denote the relative frequencies of events that fall into category $(i, j)$ based on the raw contingency table. Thus, it can be seen that the weighted average skill score can vary only based on the definition of $s_{ij}$.

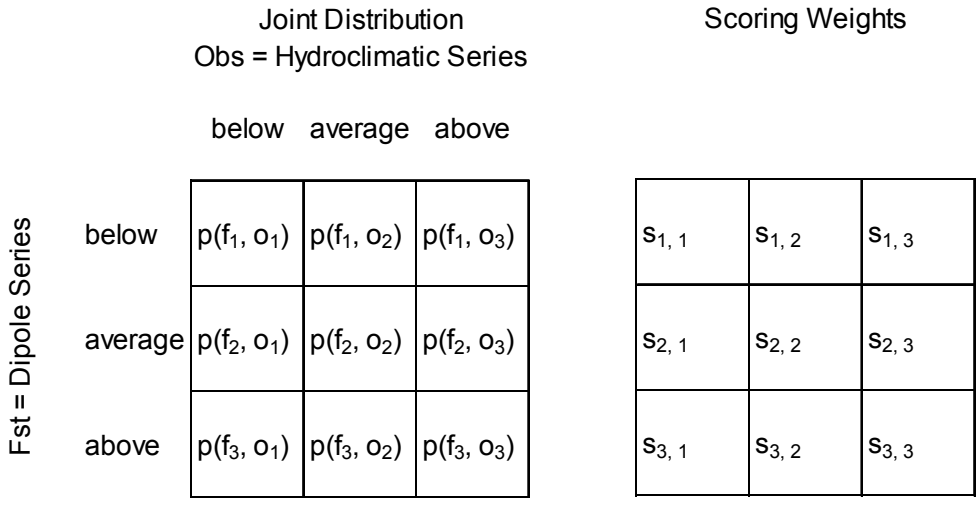



**Figure 3.3:** Performance matrix (left) and scoring matrix (right) with size $3 \times 3$.

The phrase "equitable skill score" by Gandin and Murphy (1992) is motivated by a situation in which random or constant forecasts are generated. Such forecasts are penalized by equitable weights in the off-diagonal elements of **S**. The GSS used in this study is obtained from the above equation with the particular $s_{ij}$ suggested by Gerrity (1992) as

$$s_{j,j} = \frac{1}{J-1}\left[\sum_{r=1}^{j-1}\frac{1}{D(r)} + \sum_{r=j}^{J-1} D(r)\right], j = 1,...,J; \tag{3}$$

for legitimate forecasts (diagonal elements); and

$$s_{i,j} = \frac{1}{J-1}\left[\sum_{r=1}^{j-1}\frac{1}{D(r)} + \sum_{r=j}^{J-1} D(r) - (j-i)\right], 1 \le i < j \le J; \tag{4}$$

for incorrect forecasts (off-diagonal elements). The $D(r)$ terms in both equations are the odds ratios,

$$D(j) = \frac{1 - \sum_{r=1}^{j} p(r)}{\sum_{r=1}^{j} p(r)}, j = 1,...,J-1, \tag{5}$$

based on the sample climatology $p(r)$ (i.e., the summation of each column in $\mathbf{P}$).

According to these equations, the GSS becomes zero for random forecasts (i.e., when all elements in $\mathbf{P}$ are equal) or constant forecasts (i.e., when only one row in $\mathbf{P}$ has values); otherwise, the GSS value falls within a range from -1 to 1, corresponding to the worst and perfect forecasts, respectively. Because the GSS is calculated from categorized information, it does not suffer biases even if several outliers or nonlinear trends exist in the pairwise data.

*3.2.2.2 Additional Scalar Accuracy Measures*

Depending on the number of categories, the GSS can be insensitive between various forecast combinations. To better evaluate forecasts in such cases, several other measures are discussed below and used in this work.

*Linear Correlation:* The most familiar linear correlation is Pearson's correlation estimated by

$$\rho_{f,o} = \frac{\sum_{i=1}^{h}\left(f_i - \bar{f}\right)\left(o_i - \bar{o}\right)}{\sqrt{\sum_{i=1}^{h}\left(f_i - \bar{f}\right)^2 \sum_{i=1}^{h}\left(o_i - \bar{o}\right)^2}}; \tag{6}$$

where $f$ and $o$ indicate forecasted and observed values (e.g., spring rainfall), the over-bar denotes sample means, and $h$ is the data size (e.g., the total number of years).

*Mean Absolute Error:* The mean absolute error (MAE) is given by

$$MAE = \frac{1}{h}\sum_{i=1}^{h}\left|f_i - o_i\right|, \tag{7}$$

which reflects the average forecast error magnitude. Unlike the mean squared error, which is more sensitive to large outliers, MAE weighs each data equally, reducing the overall sensitivity of the evaluation scheme.

*3.2.2.3 Ensemble Forecasting Measures*

In addition to the measures of scalar accuracy, there is a need to assess the effectiveness of ensemble forecasts. In general, such measures should be able to assess how well the forecast ensemble, composed of a forecast trace set *F*, can characterize subsequently observed values. Multiple such measures are introduced as follows. The first measure is the ensemble reliability (Re):

$$\mathrm{Re} = \frac{1}{h}\sum_{i=1}^{h} I_{F_i}\left(o_i\right), \text{ where } I_{F_i}\left(o_i\right) = \begin{cases} 1 \text{ if } o_i \in F_i, \\ 0 \text{ if } o_i \notin F_i. \end{cases} \tag{8}$$

Obviously, Re ranges between 0 (the worst) to 1 (perfect), based on how many observation points are contained by an ensemble envelope. It is possible that two forecast ensembles have the same reliability, but one suffers higher errors when violations occur. To distinguish such cases, two additional measures are introduced: The lower-minimum excess (LME),

$$LME = \underset{i=1}{\overset{h}{Min}}\left(o_i - f_i^{Min} \mid o_i < f_i^{Min}\right), \tag{9}$$

represents the largest departure of the observed values from the lower bound of the ensemble envelope. Likewise, the upper-maximum excess (UME),

$$UME = \underset{i=1}{\overset{h}{Max}}\left(o_i - f_i^{Max} \mid o_i > f_i^{Max}\right), \tag{10}$$

represents the largest departure of an observation point from the upper bound of an ensemble envelope. Lastly, the width of the ensemble forecast band, which may impact all the introduced measures, is also important. This characteristic can be expressed by the ratio ($\gamma$) of the average width of an ensemble envelope to the total observation variation:

$$\gamma = \frac{\frac{1}{h}\sum_{i=1}^{h}\left(f_i^{Max} - f_i^{Min}\right)}{\underset{i=1}{\overset{h}{Max}}\left(o_i\right) - \underset{i=1}{\overset{h}{Min}}\left(o_i\right)}. \tag{11}$$

This measure cannot be negative and has a zero value only in the case of a single forecast trace. In general, if two forecast ensembles have the same Re, LME, and UME values, the forecast ensemble with the smaller $\gamma$ value is more accurate. When $\gamma$ nears 1, the forecast ensemble skill is low and comparable to a forecast consisting of observed values. $\gamma$ values higher than 1 signify even poorer forecasts.

*3.2.3 Phased Dipole Search*

This section describes the search process for significant dipoles, including a procedure to increase computational efficiency. The need for the latter stems from the fact that the number of SST cells ( $n$ ) is usually very large leading to unmanageably high number of dipole combinations for an exhaustive search. Consider, for example, that one searches for significant dipoles consisting of one by one pixel $[1\times 1]$ [**Figure 3.2(a)**]. Such search that can be carried out over is $2\times\binom{n}{2}$ total number of dipoles (considering two-way connections). Such search is tractable even if $n$ is on the order of thousands. However, if one searches for dipoles of size $[p\times q]$ ($p$ and $q$ being cluster pixels at each pole) the number of total combinations for an exhaustive search quickly becomes intractable. To overcome this "dimensionality" problem, a systematic three phase approach is followed.

The search process begins by upscaling the original SST spatial resolution through aggregation over $(2\times 2)$ cells. If the total number of pixels in a prescribed predictor domain is $n = 500$, the overall number of combinations of $[1\times 1]$ dipoles changes from $2\times\binom{500}{2} = 249500$ to $2\times\binom{500/4}{2} = 15500$ after upscaling. In the aggregated domain, it is possible to carry out a comprehensive $[1\times 1]$ dipole search and rank the performance of dipoles according to their GSS measures.

The second phase of the search process augments the significant dipoles from size $[1\times1]$ to $[p\times q]$. Dipole augmentation proceeds by merging the existing pole regions with contiguous neighboring pixels that improve the strength of the relationship (measured by the GSS). Dipole augmentation terminates if 1) no neighboring pixel can yield improvement, or 2) the size has reached a given value *p* or *q*. **Figure 3.4(a)** illustrates how dipole augmentation (focusing on one pole) may proceed: The starting red pixel can merge with four neighboring pixels (in the four cardinal directions); after each merge, new neighboring pixels are defined, and the same procedure is repeated until termination due to either condition. More specifically, for two poles augmented simultaneously, the step-by-step procedure is followed:

Step 1: For a significant $[1\times1]$ dipole, its neighboring pixels (for both poles) are defined.

Step 2: Merging with one of its neighboring pixels one at a time, the new GSSs for all possible $[2\times1]$ or $[1\times2]$ dipoles are calculated.

Step 3: If any GSSs of $[2\times1]$ or $[1\times2]$ dipoles are higher than that of the original $[1\times1]$ dipole, the augmented dipoles along with their GSSs are recorded.

Step 4: For an augmented dipole, its neighboring pixels (for both poles) are redefined.

Step 5: Similar to Steps 2 and 3, the more complex dipoles along with their new GSSs are recorded. The process then reverts to back to step 4 for another augmented dipole.

Step 6: Steps 4 and 5 are repeated for more complex $[p \times q]$ dipoles until one of the termination conditions is met.

Step 7: Steps 1 to 6 are repeated for another significant $[1 \times 1]$ dipole until all dipoles are examined.

The second phase can construct dipoles with higher complexity and skills. However, these dipoles may not be optimal since they are still in the coarse (upscaled) spatial resolution. Thus, the third phase of the algorithm is intended to restore the original domain resolution. The significant $[p \times q]$ dipoles, thus far identified, resume a finer $[p' \times q']$ resolution and undergo local adjustments. These adjustments are carried out by the elimination and the expansion of periphery pixels until the GSS can no longer improve. The elimination procedure can be simply summarized by the steps below:

Step 1: A $[p \times q]$ dipole is first restored to its original resolution $[p' \times q']$.

Step 2: For a $[p' \times q']$ dipole, its periphery pixels (for both poles) are defined. A periphery pixel is defined as a pixel that has neighboring pixel(s) not included in the pole.

Step 3: By eliminating one of its periphery pixels one at a time, the new GSSs for all reduced $[(p'-1) \times q']$ or $[p' \times (q'-1)]$ dipoles are calculated.

Step 4: If the GSS of the $[(p'-1) \times q']$ or $[p' \times (q'-1)]$ dipole is higher than that of the original $[p' \times q']$ dipole, the reduced dipole along with its GSS are recorded.

Step 5: For a reduced dipole, its periphery pixels (for both poles) are redefined, and Steps 3 and 4 are repeated.

Step 6: Step 5 is repeated until no improvement in GSS can be made.

Step 7: Steps 1 to 6 are repeated for all significant $[p' \times q']$ dipoles.

After the dipole reduction procedure, a dipole expansion procedure, similar to the idea of the dipole augmentation in the second phase, is exerted. Although the local adjustments seem to be tedious, the experiment actually finds that the convergence rate of most identified dipoles to approach maximum skills is relatively fast. Overall, the three-phase searching process is illustrated in **Figure 3.4(b)**.

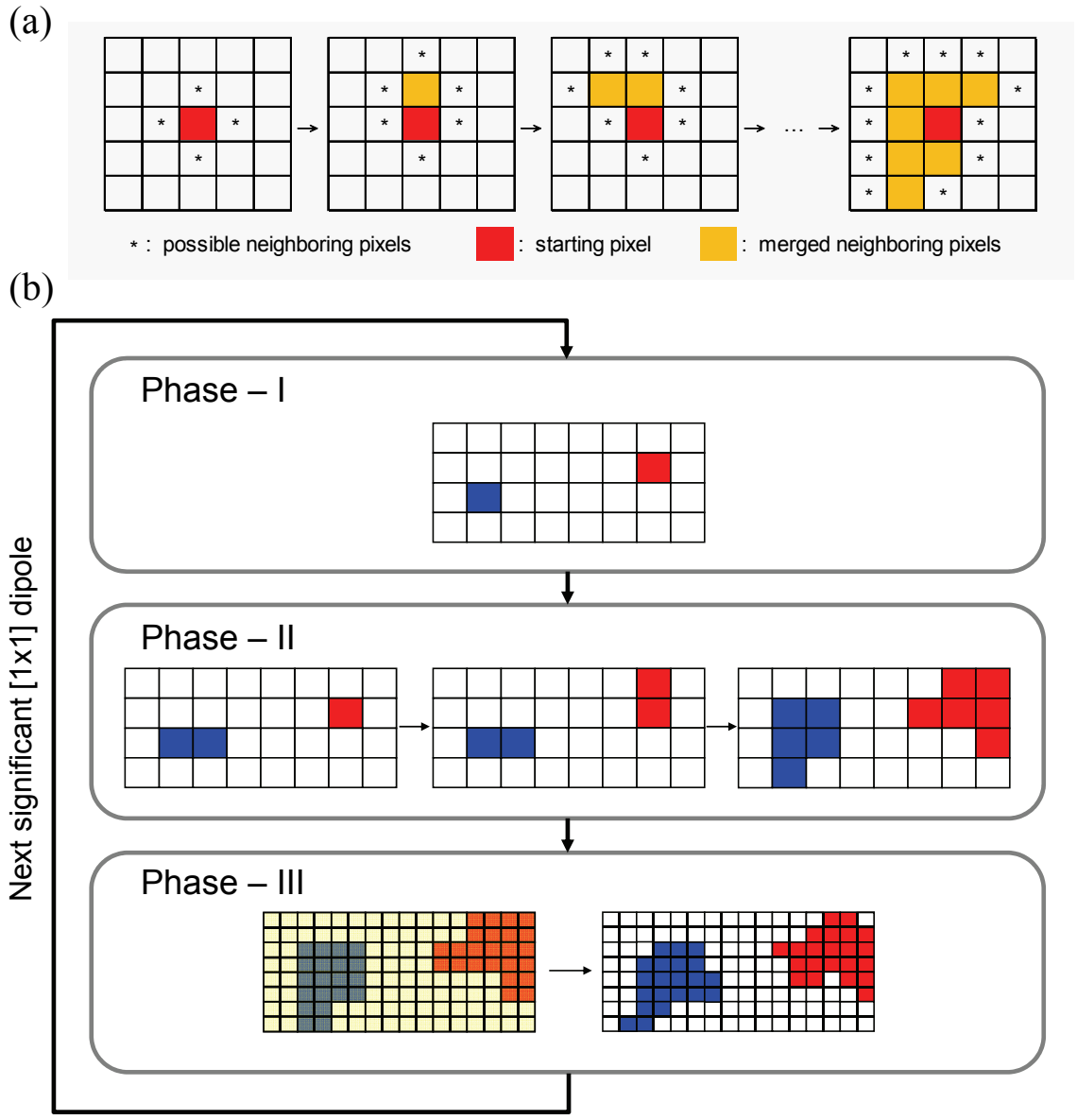



**Figure 3.4:** Formation of a dipole: (a) dipole augmentation (focusing on one pole); (b) the illustration of the three-phase dipole searching (in accordance with GSS performance).

*3.2.4 Test of Dipole Significance*

In the first phase of the search process, a GSS threshold is required to eliminate dipoles with insignificant skills indistinguishable from a random series. The significant threshold can be obtained from a specific bootstrap experiment (Efron and Tibshirani, 1993) described as follows.

For this study, the test of significance can be formulated with a null hypothesis:

$$H_0 : \ GSS_D = GSS_R, \tag{12}$$

and an alternative hypothesis:

$$H_1 : \ GSS_D > GSS_R, \tag{13}$$

where the subscripts *D* and *R* indicate that the GSS is calculated from a dipole or a random series. If $GSS_D > GSS_{R,\alpha}$ by placing a significance level (e.g., $\alpha = 0.1,\ 0.05,\ 0.025$), the null hypothesis can be rejected, and the GSS (and the associated dipole) is significant. However, to calculate $GSS_{R,\alpha}$, the distribution type of $GSS_R$ should be known. The bootstrap method provides a nonparametric and convenient means to determine this distribution.

Assuming that the size of the external observation series is $h$, a random series of the same size $h$ can be created by drawing data from the original series with replacement. Thereafter, a $GSS_R$ can be calculated between the original series and the new random series. If the same procedure is repeated for *B* random series (e.g., $B = 2000$), the distribution of $GSS_R$ as well as $GSS_{R,\alpha}$ can be obtained. **Table 3.1** summarizes the

statistics [mean ( $\mu$ ) and standard deviation ( $\sigma$ )] of $GSS_R$ calculated from four bootstrap experiments with different data sizes. Obviously, for each bootstrap experiment, the mean of $GSS_R$ is always very close to zero, indicating that a random series generally has no skill. However, the standard deviation of $GSS_R$ displays a decreasing trend from small to large data sizes. This trend is attributed to the decrease in the odds of a continuous successful forecast made by a random series when $h$ increases. For instance, if $h = 10$, the probability of a perfect forecast is $\left(1/3\right)^{10}$; and if $h = 100$, the probability of a perfect forecast is $\left(1/3\right)^{100}$. Namely, as the size of the observation series becomes large, the likelihood that a random series will generate a significant forecast decreases rapidly.

**Table 3.1:** Results of bootstrap experiments on calculating the statistics of $GSS_R$.

| $h$ | 10 | 25 | 40 | 100 |
|---|---|---|---|---|
| $\mu_{GSS_R}$ | -0.0056 | 0.0037 | -0.0024 | -0.0016 |
| $\sigma_{GSS_R}$ | 0.2380 | 0.1627 | 0.1236 | 0.0742 |

To further investigate the variability of $GSS_{R,\alpha}$, a dual-layer bootstrap can be used; that is, the aforementioned procedure ($B$ times of resampling) is repeated for another 100 (or more) times, so the distribution of $GSS_{R,\alpha}$ can be obtained. **Figure 3.5** represents the result of a dual-layer bootstrap experiment. In this experiment, the data size $(h)$ is 30, the repetition number of the inner layer $(B)$ is 2000, and the repetition number of the

outer layer is 100. The average distribution of $GSS_R$ has a quasi-symmetric bell shape with a mean value very close to zero (similar to the previous experiment), and the $GSS_{R,\alpha=0.05}$ histogram converges to a constant value ($\approx 0.25$). The results of the two experiments demonstrate that $GSS_{R,\alpha}$ is constant and only depends on data size.

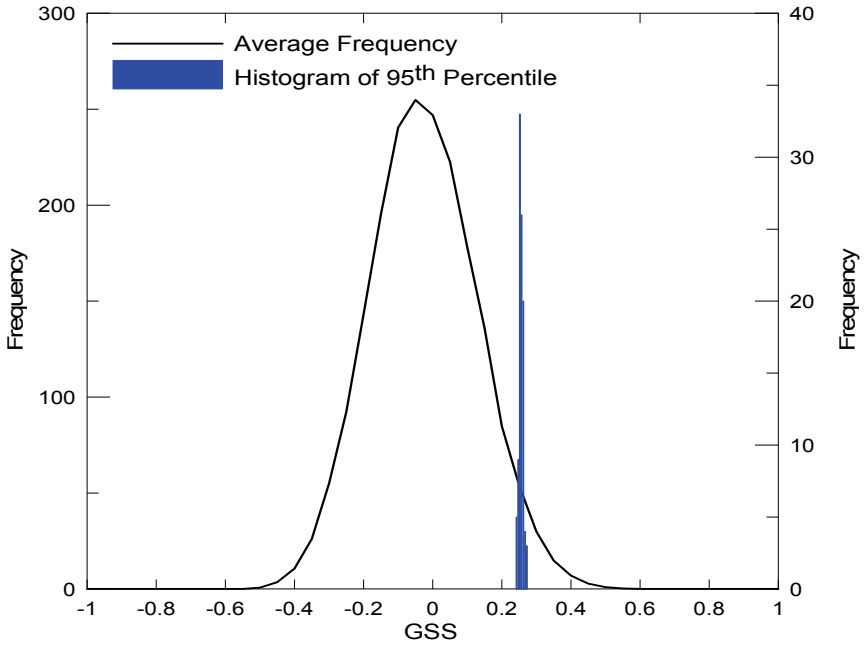



**Figure 3.5:** Bootstrap $GSS_R$ distribution along with $GSS_{R,\alpha=0.05}$ histogram.

*3.2.5 Dipole Uniqueness*

Although the three-phase search always begins with $[1\times 1]$ (base) dipoles in different locations, it is inevitable to construct identical dipoles at the end of the augmentation process. This situation is likely to occur where a specific predictor region has significant explanatory power, so any neighboring base pixels in such region may lead to

overlapping or even identical dipoles after the augmentation process. Identical dipoles may not only increase the computational burden of the algorithm but also generate duplicate forecasting results. Therefore, to ensure dipole uniqueness, an inspection scheme is applied.

According to Euclid's fundamental theorem of arithmetic (i.e., the unique-prime-factorization theorem), which states that any integer greater than one can be written as a unique product of prime numbers, the inspection procedure of dipole uniqueness is designed as follows. In a predictor domain, each pixel is assigned a prime number as its location label in addition to coordinates. When a dipole is constructed, a unique index calculated as the product of the prime numbers in the two poles is also generated. Whenever another dipole with the same index is constructed later in the search process, the algorithm eliminates it from the storing array. **Figure 3.6** illustrates this concept. As shown in the figure, the prime-number index of the first dipole (delineated by the solid boundary and connection line) can be calculated as $(47\times53\times59\times97\times101\times103\times149)\times(509\times521\times587\times593\times599\times641\times643)$. Likewise, the prime-number index of the second dipole (delineated by the dash boundary and connection line) can be calculated as $(83\times89\times97\times137\times139\times149\times193\times197\times199)\times(593\times599\times601\times647\times653\times659)$. As a consequence, although these two dipoles share some common pixels (in orange and green shading), they belong to two unique dipoles due to distinct prime-number indices.

This inspection scheme is employed two times in the three-phase searching. The first time is at the second phase when many augmented dipoles are constructed, and the second time is at the third phase to ensure no repetitive dipoles.

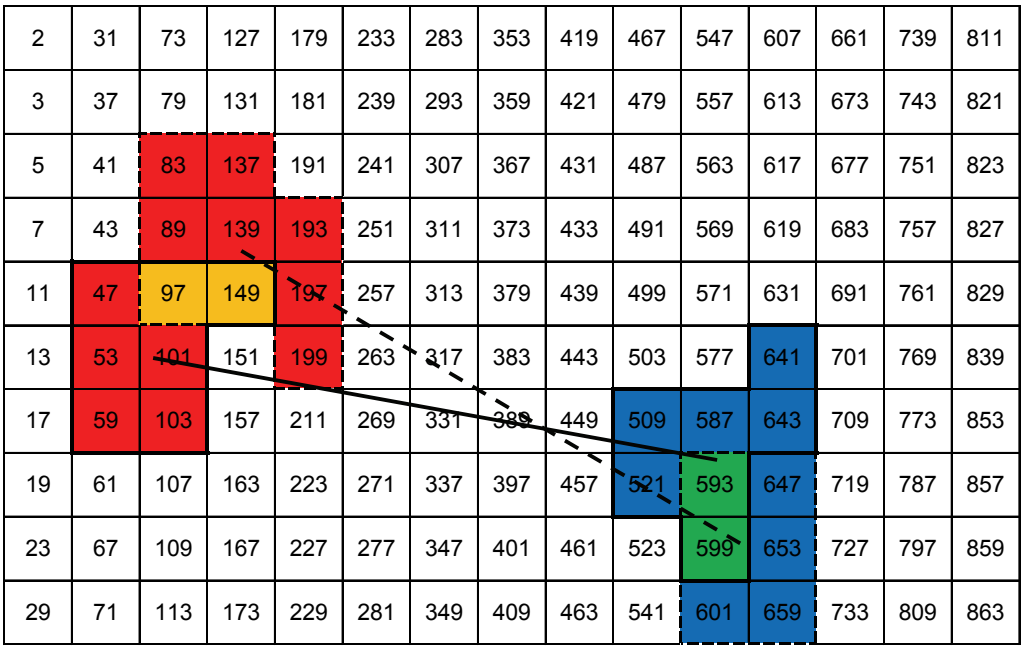



**Figure 3.6:** Illustration of inspecting dipole uniqueness using a prime number scheme. Two dipoles represented here are distinguished by solid and dash boundary lines (and connection lines). Orange and green shading indicates the overlap pixels of both poles.

## 3.3 Forecasting Process

### *3.3.1 Linear Regression*

After significant dipoles are identified by the algorithm described in the last section, these dipoles are used to generate forecasts through a predictor-predictand relationship based on linear regression. Linear regression is generally an adequate and robust tool for climate research (e.g., Folland et al., 1991; Hastenrath et al., 1995; 2004; Mutai et al., 1998), and it is applied here in the following form:

$$\mathbf{Y}^i = \mathbf{D}^i \hat{\beta}^i, \; i = 1, 2, \ldots, l, \tag{14}$$

where $\mathbf{Y}$ denotes the hydro-climatic predictand, $\mathbf{D}$ denotes the dipole predictor, $\hat{\beta}$ is the regression coefficients, the hat indicates estimated values, and $l$ is the number of significant dipoles from the final list of the three-phase search process. The coefficients $\hat{\beta}$ are estimated by the ordinary least squares. After the $l$ regression equations have been developed, $l$ hindcasting values ($\hat{\mathbf{Y}}$, i.e., retrospective forecasts) can be generated. If new observed ($\tilde{\mathbf{D}}$) values are available, the above equations can also be used to generate forecasts ($\tilde{\mathbf{Y}}$). This quasi-ensemble forecasting scheme is able to provide two important distribution moments (Kumar et al., 2000), mean and spread, revealing more information than categorical- or tercile-type forecasting.

### *3.3.2 Cross-Validation and Model Selection*

After all linear regression equations are established, each significant dipole has a GSS value (established during the dipole screening process) and a correlation coefficient ($\rho$) calculated using $\mathbf{Y}$ and $\hat{\mathbf{Y}}$. However, either of these skill measurements may be subject

to overfitting, potentially causing artificial forecasting skill. To examine this problem, typically a cross-validation (CV) scheme can be employed to extract true prediction skills from fitted models (Stone, 1974). An ordinary CV procedure such as *K*-fold CV, begins by partitioning the original dataset into *K* roughly equal-sized subsets. For the $k^{th}$ subset, the model is first fitted based on the other $k-1$ subsets (calibration set), and then the forecasting error of the fitted model is calculated based on the $k^{th}$ subset (validation set). The partitioning procedure is repeated to compute the expectation of forecast errors until every subset is evaluated. If the number *K* is equal to data size *h*, this procedure is known as "leave-one-out cross-validation." The model with the smallest expectation of forecasting errors (e.g., the smallest MAE) is finally selected. A more general discussion on CV procedures applied to climate forecasting practices can be found in Michaelsen (1987) and Elsner and Schmertmann (1994).

In the context of this study, these similar concepts are customized as follows. Each significant dipole identified by the algorithm goes through the leave-one-out CV procedure by repeatedly fitting $h$ regression equations by reserving one data point at a time (starting with the first) and using the fitted equation to estimate the reserved data point. This yields $\hat{y}_i^{CV}$, $i = 1, 2, \ldots, h$. After obtaining *h* estimated values, the CV MAE for this dipole can be calculated using the following equation:

$$MAE_{CV} = \frac{1}{h}\sum_{i=1}^{h}\left|\hat{y}_i^{CV} - y_i\right|. \tag{15}$$

Since *l* regressive equations have been established, the above procedure has to be repeated *l* times to calculate $MAE_{CV}$ for all dipoles. To create ensemble forecasts, the

top $\xi$ dipoles (e.g., $\xi = 10$ or $20$) with the lowest $MAE_{CV}$ values are selected. The schematic flow chart of this leave-one-out CV procedure is illustrated in **Figure 3.7**.

Furthermore, to detect and eliminate possible multi-collinearity issues resulting from the selection of highly overlapping dipoles or highly correlated dipole time series, the variance inflation factor (VIF) is computed:

$$VIF_j = \frac{1}{1 - R_j^2}, \tag{16}$$

where $R_j^2$ is the coefficient of determination obtained from a regression of the $j^{th}$ dipole time series on any other dipole time series. In typical regression analysis, a VIF value larger than 10 indicates that multi-collinearity exists among predictors (Neter et al., 1996).

To established the multi-collinearity threshold, the variations of Re as well as $\gamma$ can be assessed from experiments that adjust different VIF values as criteria in the dipole selection process. Two experiments based on the dipole identification for the winter and spring precipitation in the southeastern United States (a full case study will be carried out in Chapter 4) present the variations of Re and $\gamma$ in **Figure 3.8**. Bothe experiments reveal similar results. While $\gamma$ seems to decrease monotonously when the VIF changes from low to high values, Re has a possible maximum when the VIF ranges between 2 to 4. A high VIF value (~ 10) indicates that most of the selected dipoles are highly correlated thereby reducing both $\gamma$ and Re. On the other hand, a very low VIF value (< 2) that overemphasizes the distinction among the selected dipoles may sacrifice their forecasting accuracy ($MAE_{CV}$ and $\rho_{CV}$). As a result, a wider but flatter forecasting envelope is

produced, reducing Re. According to these results, a VIF value of 4 represents a reasonable compromise and is adopted in the following case studies.

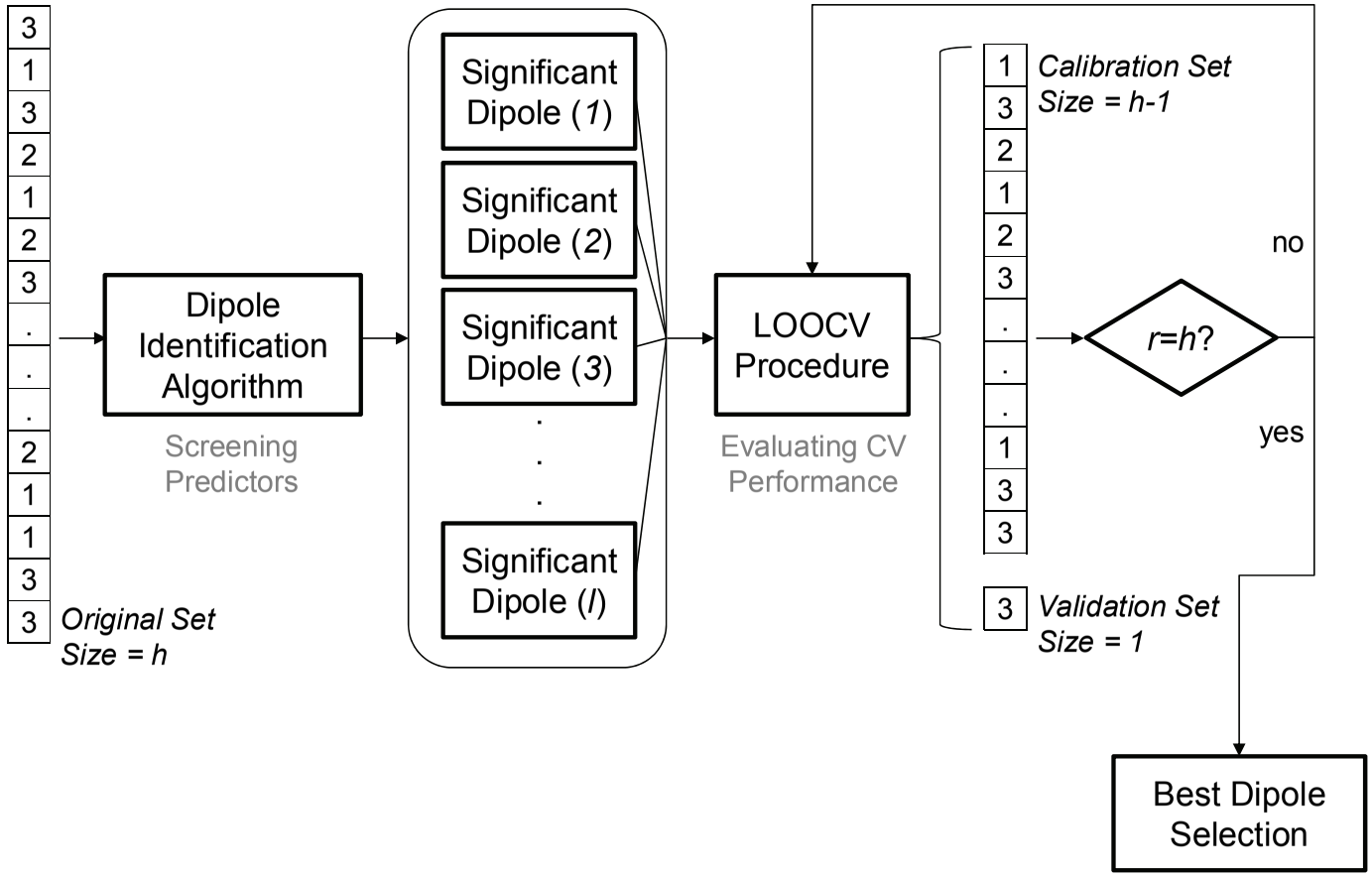



**Figure 3.7:** Schematic flow chart of the dipole identification algorithm along with the leave-one-out CV procedure.

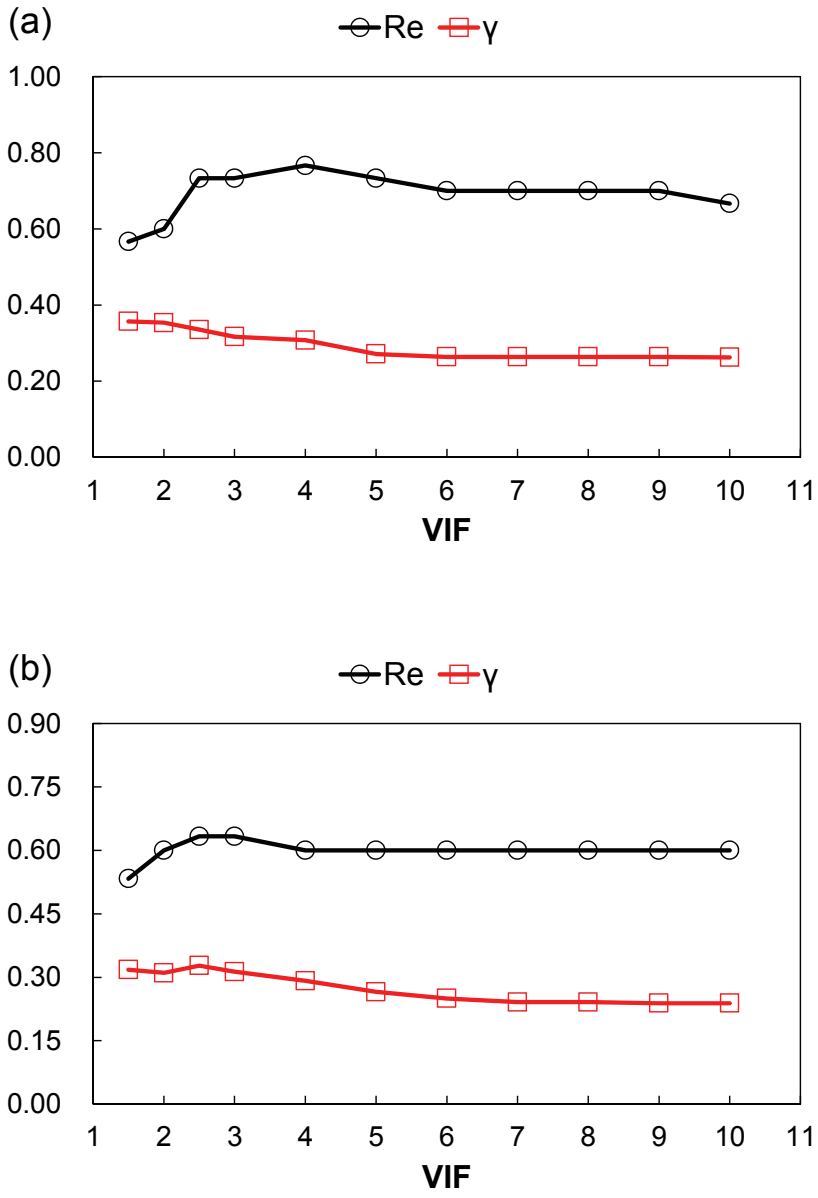



**Figure 3.8:** Variations of two ensemble forecasting measures (Re and $\gamma$) corresponding to the change in VIF.

# CHAPTER 4

# CASE STUDY IN THE SOUTHEAST US

This chapter applies the method described in Chapter 3 to the forecasting of seasonal precipitation in a major river basin in the southeastern US. Section 4.1 describes the background information of this river basin. Section 4.2 first presents the general method set-up and then interprets associated results including the synopses of forecasting skills and the illustrations of identified dipoles. Section 4.3 compares the results generated by the dipole method with those generated by other statistical methods. Section 4.4 performs climate diagnostics based on composite analysis to draw some possible connections of the identified dipoles with the upper-level and surface variables and then discusses plausible underlying physical mechanisms. Section 4.5 demonstrates an operational forecasting framework and shows the latest forecasting results for 2012 seasonal precipitation.

## 4.1 Southeast Seasonal Precipitation

One of the major river basins in the southeastern US is the Apalachicola-Chattahoochee-Flint (ACF) River Basin shown in **Figure 4.1**. The ACF basin has a drainage area of 19,300 square miles across Georgia, Alabama, and Florida, where it serves many water resources interests. Reliable seasonal forecasting would benefit the ACF management, especially during droughts when demands exceed supplies. This end, the Flint River Drought Protection Act requires that on March 1 each year, the Georgia State Government must determine and declare whether a severe drought is anticipated for the

coming months. If a drought is anticipated, various water withdrawal reductions are mandated, including over the agricultural and heavily irrigated areas around the Flint River.

The climate of the ACF basin varies from the upper to the lower reaches. Above the fall line, the upper ACF (UACF) basin includes the Buford, West Point, and Montezuma sub-basins, while below the fall line, the lower ACF (LACF) basin includes the George, Albany, and Bainbridge sub-basins.

**Figure 4.1** also presents the monthly precipitation climatology of these two regions derived from the PRISM dataset. Generally, the two most significant rainy seasons in the Southeast are winter (December–February) and summer (June–August). The climatology series of the UACF and LACF are highly correlated, but they also demonstrate different climatic influences. From fall to spring, UACF usually receives more precipitation than LACF because of the southward intrusion of extratropical cyclonic systems and the local orographic effect in northeastern Georgia. In summer, in contrast, LACF usually receives more precipitation than UACF because of active convective storm systems originating from the Gulf of Mexico and the Atlantic Ocean. Thus, in the analysis that follows, the two series are investigated separately.

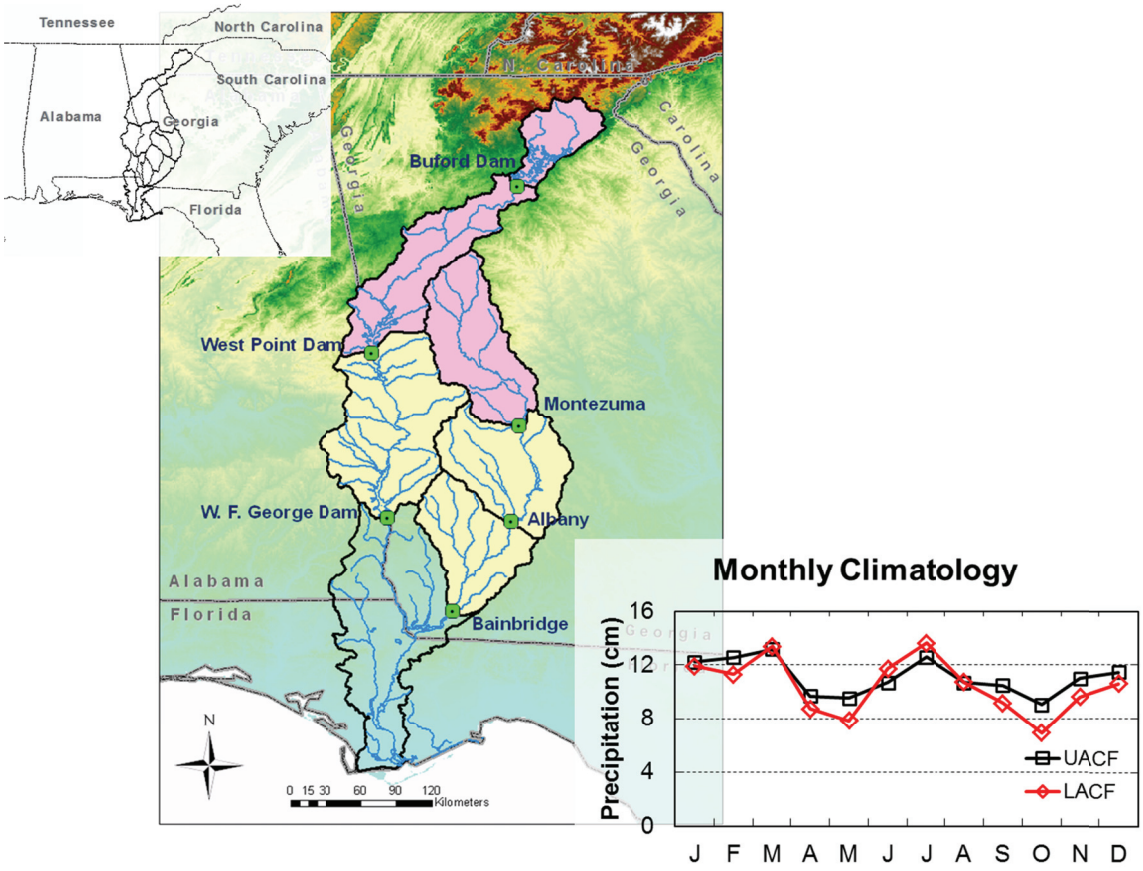



**Figure 4.1:** The ACF basin in the southeastern US and corresponding monthly precipitation climatology. The upper ACF (UACF) basin comprises Buford, West Point, and Montezuma sub-basins (in rose color), and the lower ACF (LACF) basin comprises George, Albany, and Bainbridge sub-basins (in yellow color).

## 4.2 Seasonal Dipole Identification

In this section, the dipole algorithm is applied to the hindcasting (retrospective forecasting experiment) of seasonal precipitation. Precipitation series in the four seasons—DJF, MAM, JJA, and SON—from 1981 to 2010 for both UACF and LACF are generated as the target external series for forecasting. The purpose of this study is to examine whether the proposed algorithm can identify significant SST-dipole predictors that have skills to explain the ACF precipitation in the four seasons.

The SST data used for this study is the Kaplan Extended SST V2 product as introduced in Chapter 3. **Figure 4.2** shows a defined SSTA spatial domain (10°S–70°N, 180°W–0°) that covers the major ENSO region (the eastern tropical Pacific), the tropical Atlantic, a large portion of the North Pacific, and nearly the entire North Atlantic. Because this SST domain, which encompasses North America, closely interacts with atmospheric systems influencing the North American climate, it is presumed that the most sensitive predictors governing the southeastern US climate should also reside in this domain.

Furthermore, to explore the most skillful lead time for forecasting, not only the dipole algorithm is executed in every concurrent season of external series, but also the process is repeated for twelve lead times (corresponding to 1, 2, …, 12 months) for both UACF and LACF. These and other user-specified parameters are summarized in **Table 4.1**.

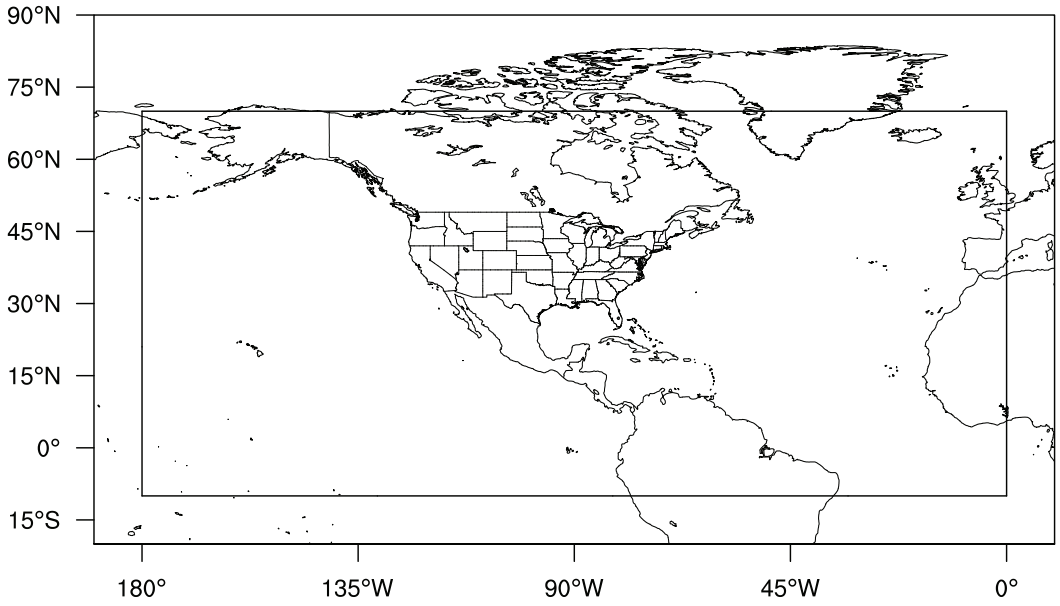



**Figure 4.2:** SSTA predictor domain used for the case study in the Southeast US.

**Table 4.1:** User-specified parameters for the case study in the Southeast US.

| **Data size** $(h)$ | **Predictor description** | **Predictand description** | **Upper threshold of clustering pixels** $[p \times q]$ | $GSS_{R,\alpha=0.05}$ | $\xi$ | **Lead times** |
|---|---|---|---|---|---|---|
| 30 (1981–2010; DJF, MAM, JJA, and SON) | Kaplan SSTA [10°S–70°N, 180°W–0°] | Prism-derived seasonal precipitation for UACF and LACF | $[p = q = 10]$ | 0.25 | 20 | 0, 1, 2, …, 12 (months) |

*4.2.1 Forecasting Skill Summary*

The dipole algorithm is applied to identify any significant dipoles that for a given lead time have significant forecasting skill. The forecasting skill is characterized by two scalar accuracy measures (MAE and $\rho$) and four ensemble forecasting measures (Re, LME, UME, and $\gamma$) as recorded in **Tables 4.2** and **4.3** for UACF and LACF, respectively.

In general, the strongest hindcasting performance appears to be in DJF for both UACF and LACF, indicating that oceanic temperatures play a dominant role in driving the winter regional climate. When the lead time is three months (i.e., when SON SSTs are used to predict DJF precipitation), the dipole ensemble forecast has the highest Re (0.833), and the best single dipole trace also has the lowest MAE (1.542) for UACF. Another skillful lead time for UACF occurs in the eighth month (i.e., AMJ SST dipole predictors), producing the best LME (-0.901) and UME (1.641) as well as the best single dipole trace that can explain about 57% of the DJF precipitation variance. $\gamma$ values also vary in a reasonable range from 33% to 42% throughout all lead times. Regarding the DJF forecasting results for LACF, the most skillful lead time is found at zero months with high $\rho$ (0.729) and Re (0.767) values. This reflects a strong concurrent relationship between the SST field and the winter LACF precipitation. However, the dipole forecasts based on the 4- and 9-month lead times (with real forecasting implications) seem to still exhibit some credible skills (e.g., $\rho = 0.670$ and 0.723; Re = 0.733 and 0.667 ). Coincidentally, the optimum lead times identified for UACF and LACF are somewhat consistent, but the corresponding dipoles show slightly higher skill for UACF.

In contrast to the very strong predictability seen for the DJF precipitation, the MAM precipitation can only be predicted with moderate to weak skill. The ACF precipitation during this climatic transition season seems to be driven much by internal atmospheric variability, so the dipole algorithm has difficulty in extracting strong oceanic predictors. Even though the dipole algorithm sifts through all lead times, the dipole prediction skill is not as strong as those for DJF. The optimum lead time is found to be twelve months for both UACF and LACF in terms of the Re values (0.667 for the former and 0.633 for the latter). The best single dipole trace can account for 43% of precipitation variance for UACF but only 24% for LACF.

The JJA hindcasting results show moderate skills. Higher Re values (0.600) can be seen at multiple lead times for UACF. If other skill measures are examined together, dipoles at the 4- and 8- to 10-month lead times seem to have better integrated performance. The best single dipole trace at the 4-month lead time can explain about 32% of the JJA precipitation variance. At the same lead time, dipoles identified for the JJA precipitation in LACF have improved forecasting skills (e.g., $\rho = 0.604$ and $\mathrm{Re} = 0.700$). Another forecasting skill peak appears to be at the 7–9-month lead times similar to those found for UACF. Notably, the UME values in this season are unreasonably high in comparison with those in other seasons. Such high UME values are associated with unpredictable hurricane/ tropical storm events (e.g., Tropical Storm Alberto in 1994 and Hurricane Katrina in 2005). The underestimation of this kind of intense precipitation is not surprising because the forecasting method based on simple linear regression may have difficulty reflecting accurate amplitude resulting from more “nonlinear” tropical

disturbances. If the skill measures are reevaluated by excluding the two aforementioned hurricane/ tropical storm events, the updated results (at the 4-month lead time) can be shown in **Table 4.4**. Almost all measures exhibit substantial improvement (~15% in MAE, 7% in Re, 33 and 63% in UME) except $\gamma$ which increases due to the reduced total data variation.

Lastly, the SON hindcasting results for both UACF and LACF display the second strongest skill in the four seasons. Unlike the spring, the fall as the other climatic transition season seems to have more precipitation events driven by synoptic-scale systems so that they correlate better with teleconnection patterns. As a result, when the lead times are three to five (i.e., SON precipitation is predicted by AMJ to JJA SSTs), the dipole ensemble forecasts can reach a Re value as high as 0.700, and the best single dipole trace can explain about 40% of precipitation variance for both UACF and LACF. However, following the increased lead times, a fast-decaying rate of forecasting skills is observed. In addition, similar to the summer tropical disturbances, some hurricane/tropical storm events (e.g., Tropical Storm Hanna in 2002 and Hurricane Ivan in 2004) also affect the SON precipitation. Therefore, the same reevaluation procedure (as for the JJA precipitation) is applied, and the updated forecasting skills (at the 3-month lead time) are shown in **Table 4.4**. As before, most of the skill measures improve, except UME which remains the same. This is probably due to a hurricane-unrelated discrepancy between the observation and the hindcasting value in 2009. The 2009 devastating Southeast flood was actually attributed to a rare favorable mixing of both synoptic and mesoscale conditions (Shepherd et al., 2011).

Although the dipole method has shown the prominent skills for almost all seasonal precipitation in the ACF basin through the hindcasting experiment, some questions remain: 1) Why do the skill measure occasionally show multiple optima at different lead times? 2) How can an operational forecasting framework be established based on these hindcasting results? A possible answer to the first question is that the hindcasting experiment does not utilize enough explanatory variables. That is, if other hidden explanatory variables in addition to the specified SSTA field account for some precipitation variance, they may be able to fill up the gaps in the forecasting skills and then present a smooth skill transition across the lead times. The second question will be addressed in the last section of this chapter.

**Table 4.2:** Forecasting skill at different lead times for UACF; the best value of each measure is bold and underlined.

| UACF_DJF | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 0 | 1.729 | 0.672 | 0.600 | -3.178 | 3.478 | **0.333** |
| 1 | 1.800 | 0.622 | 0.700 | -3.431 | 3.413 | 0.358 |
| 2 | 1.659 | 0.672 | 0.667 | -2.708 | 3.105 | 0.392 |
| 3 | **1.542** | 0.725 | **0.833** | -4.153 | 2.309 | 0.371 |
| 4 | 1.656 | 0.683 | 0.600 | -4.745 | 3.214 | 0.350 |
| 5 | 1.786 | 0.658 | 0.700 | -4.779 | 2.646 | 0.363 |
| 6 | 1.728 | 0.666 | 0.733 | -3.662 | 1.986 | 0.418 |
| 7 | 1.707 | 0.702 | 0.633 | -2.185 | 2.240 | 0.391 |
| 8 | 1.625 | **0.757** | 0.733 | **-0.901** | **1.641** | 0.409 |
| 9 | 1.654 | 0.715 | 0.667 | -1.214 | 2.106 | 0.377 |
| 10 | 1.903 | 0.517 | 0.567 | -4.622 | 3.455 | 0.338 |
| 11 | 2.064 | 0.460 | 0.600 | -3.738 | 3.348 | 0.348 |
| 12 | 2.000 | 0.567 | 0.667 | -3.546 | 3.086 | 0.387 |
| **UACF_MAM** | | | | | | |
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 0 | 2.186 | 0.368 | 0.533 | **-3.435** | **3.618** | 0.400 |
| 1 | 2.344 | 0.412 | 0.467 | -4.142 | 4.870 | 0.381 |
| 2 | 2.314 | 0.418 | 0.433 | -4.981 | 4.913 | 0.301 |
| 3 | 2.355 | 0.462 | 0.433 | -4.866 | 5.412 | 0.321 |
| 4 | 2.289 | 0.423 | 0.433 | -5.018 | 5.420 | 0.322 |
| 5 | 2.364 | 0.412 | 0.433 | -4.905 | 5.806 | 0.308 |
| 6 | 2.436 | 0.351 | 0.400 | -3.791 | 4.374 | 0.323 |
| 7 | 2.425 | 0.270 | 0.400 | -4.511 | 5.092 | 0.311 |
| 8 | 2.393 | 0.385 | 0.433 | -4.235 | 5.269 | 0.302 |
| 9 | 2.332 | 0.343 | 0.333 | -5.223 | 5.457 | **0.253** |
| 10 | 2.280 | 0.401 | 0.500 | -5.054 | 3.970 | 0.333 |
| 11 | 2.089 | 0.520 | 0.567 | -4.448 | 3.805 | 0.372 |
| 12 | **1.913** | **0.653** | **0.667** | -3.944 | 4.358 | 0.404 |

**Table 4.2:** *Continued*.

| UACF_JJA | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 0 | 2.501 | 0.340 | 0.500 | -3.178 | 7.463 | **0.313** |
| 1 | 2.463 | 0.335 | **0.600** | -2.877 | 7.703 | 0.326 |
| 2 | 2.359 | 0.486 | 0.500 | -2.904 | **5.653** | 0.355 |
| 3 | 2.374 | 0.551 | 0.500 | -1.741 | 7.174 | 0.340 |
| 4 | 2.336 | **0.566** | **0.600** | -2.582 | 6.327 | 0.392 |
| 5 | 2.372 | 0.391 | 0.567 | -2.365 | 6.675 | 0.356 |
| 6 | 2.476 | 0.500 | **0.600** | -2.369 | 5.772 | 0.343 |
| 7 | 2.459 | 0.417 | 0.467 | -3.379 | 6.295 | 0.322 |
| 8 | 2.346 | 0.551 | 0.567 | -1.359 | 6.410 | 0.370 |
| 9 | 2.485 | 0.461 | **0.600** | **-1.178** | 5.949 | 0.375 |
| 10 | 2.399 | 0.542 | **0.600** | -1.657 | 7.014 | 0.360 |
| 11 | 2.374 | 0.512 | 0.500 | -1.986 | 6.756 | 0.358 |
| 12 | **2.318** | 0.558 | 0.500 | -3.094 | 5.676 | 0.359 |
| **UACF_SON** | | | | | | |
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 0 | 2.786 | 0.425 | 0.467 | -3.403 | 7.075 | 0.323 |
| 1 | 2.663 | 0.509 | 0.500 | -3.510 | 8.295 | 0.336 |
| 2 | 2.682 | 0.497 | 0.567 | -2.427 | 7.373 | 0.377 |
| 3 | 2.613 | 0.542 | **0.700** | **-1.968** | 7.756 | 0.398 |
| 4 | 2.472 | 0.537 | 0.533 | -2.568 | 6.667 | 0.386 |
| 5 | **2.458** | **0.612** | 0.633 | -2.525 | **5.945** | 0.362 |
| 6 | 2.689 | 0.475 | 0.500 | -3.080 | 8.521 | 0.344 |
| 7 | 2.932 | 0.472 | 0.400 | -4.045 | 8.651 | 0.294 |
| 8 | 2.726 | 0.460 | 0.433 | -4.060 | 8.839 | 0.311 |
| 9 | 3.345 | -0.010 | 0.333 | -4.152 | 9.218 | **0.242** |
| 10 | 2.908 | 0.344 | 0.367 | -4.278 | 9.322 | 0.254 |
| 11 | 2.952 | 0.302 | 0.400 | -3.591 | 9.001 | 0.284 |
| 12 | 3.018 | 0.390 | 0.433 | -3.067 | 7.817 | 0.322 |

**Table 4.3:** As in **Table 4.2**, but for LACF.

| LACF_DJF | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 0 | 1.683 | **0.729** | **0.767** | -2.454 | 4.318 | 0.358 |
| 1 | 1.854 | 0.659 | 0.667 | -2.282 | 5.222 | 0.371 |
| 2 | 1.797 | 0.655 | 0.700 | -1.774 | 5.091 | 0.385 |
| 3 | 1.742 | 0.673 | 0.667 | -2.846 | 5.606 | 0.366 |
| 4 | 1.734 | 0.670 | 0.733 | -2.344 | 5.475 | 0.337 |
| 5 | 1.839 | 0.627 | 0.433 | -2.823 | 5.501 | 0.336 |
| 6 | 1.934 | 0.587 | 0.567 | -2.486 | 5.861 | 0.345 |
| 7 | 1.779 | 0.690 | 0.633 | -2.592 | 5.043 | 0.359 |
| 8 | 1.742 | 0.615 | 0.667 | -2.814 | 4.299 | 0.361 |
| 9 | **1.563** | 0.723 | 0.667 | **-1.572** | 4.936 | 0.341 |
| 10 | 1.761 | 0.695 | 0.667 | -2.471 | 5.214 | 0.333 |
| 11 | 1.873 | 0.584 | 0.533 | -2.154 | 4.215 | **0.320** |
| 12 | 1.898 | 0.693 | 0.467 | -2.405 | **3.496** | 0.321 |
| **LACF_MAM** | | | | | | |
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 0 | 2.390 | 0.367 | 0.467 | -3.888 | 6.520 | 0.312 |
| 1 | 2.586 | 0.273 | 0.400 | -3.913 | 7.302 | 0.275 |
| 2 | 2.494 | 0.333 | 0.433 | -3.994 | 6.616 | **0.246** |
| 3 | 2.350 | **0.548** | 0.500 | -4.491 | 5.095 | 0.277 |
| 4 | 2.399 | 0.444 | 0.500 | -4.571 | 5.938 | 0.264 |
| 5 | 2.462 | 0.380 | 0.467 | -4.271 | 6.608 | 0.283 |
| 6 | 2.477 | 0.257 | 0.500 | -3.847 | 7.028 | 0.278 |
| 7 | 2.540 | 0.237 | 0.433 | -4.260 | 7.017 | 0.257 |
| 8 | 2.592 | 0.076 | 0.400 | -4.344 | 7.262 | 0.256 |
| 9 | 2.629 | 0.344 | 0.433 | -4.377 | 6.730 | 0.254 |
| 10 | 2.616 | 0.278 | 0.367 | -4.351 | 5.651 | 0.307 |
| 11 | 2.489 | 0.240 | 0.533 | **-3.134** | 6.246 | 0.354 |
| 12 | **2.234** | 0.486 | **0.633** | -3.777 | **4.827** | 0.391 |

**Table 4.3:** *Continued*.

| LACF_JJA | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 0 | 2.374 | 0.213 | 0.433 | -3.150 | 8.444 | **0.232** |
| 1 | 2.335 | 0.304 | 0.467 | -2.637 | 8.738 | 0.275 |
| 2 | 2.305 | 0.360 | 0.600 | -2.801 | 8.668 | 0.276 |
| 3 | **2.043** | 0.555 | 0.667 | -2.727 | 8.939 | 0.293 |
| 4 | 2.146 | **0.604** | **0.700** | -2.687 | 8.040 | 0.338 |
| 5 | 2.153 | 0.456 | 0.633 | -2.633 | 8.290 | 0.327 |
| 6 | 2.346 | 0.369 | 0.600 | -2.035 | 8.847 | 0.299 |
| 7 | 2.424 | 0.418 | 0.633 | -2.265 | 8.827 | 0.306 |
| 8 | 2.219 | 0.519 | 0.567 | -1.814 | **7.386** | 0.339 |
| 9 | 2.358 | 0.345 | 0.567 | **-1.793** | 7.479 | 0.344 |
| 10 | 2.445 | 0.292 | 0.533 | -2.358 | 8.944 | 0.304 |
| 11 | 2.328 | 0.348 | 0.500 | -2.730 | 8.467 | 0.318 |
| 12 | 2.322 | 0.443 | 0.567 | -1.815 | 8.137 | 0.319 |
| **LACF_SON** | | | | | | |
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 0 | 2.119 | 0.481 | 0.500 | -4.011 | 2.513 | 0.381 |
| 1 | 2.078 | 0.579 | 0.600 | -4.139 | **1.846** | 0.378 |
| 2 | 1.900 | 0.593 | 0.567 | -3.997 | 2.482 | 0.359 |
| 3 | 1.872 | 0.623 | **0.700** | **-2.241** | 3.561 | 0.411 |
| 4 | **1.770** | **0.643** | 0.667 | -2.825 | 2.965 | 0.420 |
| 5 | 1.941 | 0.522 | 0.600 | -3.523 | 4.443 | 0.363 |
| 6 | 2.275 | 0.252 | 0.433 | -4.697 | 5.028 | 0.302 |
| 7 | 2.221 | 0.382 | 0.333 | -4.726 | 5.117 | 0.319 |
| 8 | 2.347 | 0.266 | 0.367 | -3.124 | 5.365 | 0.309 |
| 9 | 2.443 | 0.457 | 0.400 | -3.190 | 5.444 | **0.295** |
| 10 | 2.285 | 0.388 | 0.433 | -4.425 | 5.496 | 0.326 |
| 11 | 2.243 | 0.478 | 0.467 | -2.897 | 5.367 | 0.360 |
| 12 | 2.420 | 0.338 | 0.467 | -3.685 | 4.833 | 0.377 |

**Table 4.4:** Updated forecasting skill for JJA and SON precipitation at selected lead times (numbers in parentheses) after excluding two influential hurricane/tropical storm (-2 HS) events in each season.

| UACF | MAE | ρ | Re | LME | UME | γ |
|---|---|---|---|---|---|---|
| JJA (4) RAW | 2.336 | 0.566 | 0.600 | -2.582 | 6.327 | 0.392 |
| JJA (4) -2 HS | 1.993 | 0.571 | 0.643 | -2.582 | 4.230 | 0.477 |
| SON (3) RAW | 2.613 | 0.542 | 0.700 | -1.968 | 7.756 | 0.398 |
| SON (3) -2 HS | 2.445 | 0.531 | 0.750 | -1.968 | 7.756 | 0.405 |
| **LACF** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| JJA (4) RAW | 2.146 | 0.604 | 0.700 | -2.687 | 8.040 | 0.338 |
| JJA (4) -2 HS | 1.844 | 0.608 | 0.750 | -2.687 | 2.952 | 0.472 |
| SON (3) RAW | 1.872 | 0.623 | 0.700 | -2.241 | 3.561 | 0.411 |
| SON (3) -2 HS | 1.640 | 0.702 | 0.750 | -2.241 | 1.934 | 0.419 |

*4.2.2 Identified Dipole Illustrations*

To illustrate the actual variations of the ensemble forecasting and observation values, the optimal hindcasting results (time series) at two different lead times for each season are selected and plotted in **Figures 4.3** and **4.4** (UACF) and **Figures 4.7** and **4.8** (LACF). The selection of lead times for result illustrations basically uses the highest Re values and some additional criteria: 1) Only lead times longer than or equal to three months are selected (for forecasting implications); 2) Two selected lead times have as fewer overlapping periods as possible (for more distinct patterns); and 3) Two selected lead times for UACF are the same or very similar to those for LACF (for a consistent comparison). Furthermore, the corresponding dipoles used to generate these hindcasting results are also plotted in **Figures 4.5** and **4.6** (UACF) and **Figures 4.9** and **4.10** (LACF).

**Figure 4.3(a)** presents the hindcasting result at the first selected lead time (three months) for the DJF precipitation in UACF. Because of a very low MAE value and very high $\rho$ and Re values, the figure reveals only minor discrepancy between the observation and the best forecasting traces as well as a reliable forecasting envelop that contains most of the observations. The corresponding dipoles [**Figure 4.5(a)**] that generate this hindcasting result are mainly constituted by a positive pole region along the West to Northwest Coast of North America and a negative pole region from the Labrador Sea to the coast of Iceland. This positive pole implies some connections to ENSO since this warming region is an extension of El Niño due to the Kelvin wave mechanism (Johnson and O'Brien, 1990), which has been discussed in many studies (e.g., Lau and Nath, 1996; Alexander et. al., 2002). **Figure 4.3(b)** presents the other hindcasting result at the 8-month lead time. Although the Re value decreases somewhat, the overall correspondence between the observations and the forecasting envelop is exceptional, leading to even better LME and UME values. The corresponding dipoles [**Figure 4.5(b)**] display some changes in the locations, but some similar elements and orientations are reserved. While the positive poles are away from the West Coast of North America and more extensive, the main negative poles are separately located in the coast of Iceland and the western coast of Mexico.

The hindcasting results at two selected lead times (four and nine months) for LACF [**Figures 4.7(a) and 4.7(b)**] behave similarly in comparison with those for UACF. A higher Re value is observed for the results at the short lead time, but better LME and UME values are found for the results at the long lead time. However, the corresponding

dipoles at the two lead times for LACF seem to be organized differently from those for UACF. **Figure 4.9(a)** shows that the dipoles at the short lead time consist of a more extensive positive pole region over the tropical to subtropical Pacific Ocean (related to ENSO) and a more concentrated negative pole region in the vicinity of Iceland. More differences between the dipoles for UACF and those for LACF can be seen at the long lead time. **Figure 4.9(b)** shows that the only active region for positive poles is situated at the North Pacific, and the active regions for negative poles are scattered over the Atlantic Ocean (and some over the Pacific). Despite these differences, these dipole patterns for both UACF and LACF (especially at the short lead time) indicate that ENSO is of great importance to the DJF precipitation (especially for LACF since there are more tropical signals). Thus, the results reconfirm that the warm (cold) phase of ENSO may increase (decrease) winter precipitation in the southeastern US, which has been addressed in several studies (e.g., Ropelewski and Halpert, 1986; Kurtzman and Scanlon, 2007; Mo and Schemm, 2008; Seager et al., 2009). However, this predictor-predictand relationship can further be enhanced by introducing the dipole structure with specific spatial configurations. Furthermore, more sensitive dipole regions identified by the algorithm may provide a clue that the variation of the DJF precipitation is linked to more essential phenomena (e.g., jet streams) beyond the ENSO impacts.

**Figures 4.3(c)** and **4.3(d)** present the hindcasting results at two selected lead times (twelve and three months) for the MAM precipitation in UACF. The result at the long lead time is apparently better in all skill measures, but the associated skills seem to deteriorate after 1999. The corresponding dipoles at the long lead time [**Figure 4.5(c)**]

show two active regions: positive poles at the coast of Western Europe and negative poles at the coast of Iceland. Because of weaker forecasting skills, the corresponding dipoles at the short lead time [**Figure 4.5(d)**] show only one concentrated positive pole region situated at the western coast of Mexico to California. Other dipoles/poles are basically scattered all over the oceanic areas, which is also another indicator of insignificant forecasting skills. Regarding the forecasts for LACF [**Figures 4.7(c)** and **4.7(d)]**, they also respond to the observations in a very similar manner (better skills for the result at the long lead time with some deterioration after 1999), but the deterioration of skills is more considerable. Furthermore, the corresponding dipoles at the two lead times for LACF [**Figures 4.9(c)** and **4.9(d)]** also share very similar active regions; however, the positive poles at the short lead time seem to have some southward displacement and divide into two regions. Such regions may exhibit the implication for guiding subtropical jet streams, but more related work is required to support this argument. Apart from the patterns described above, other systematic patterns for the MAM precipitation are difficult to observe and interpret. Moreover, the deterioration of skills after 1999 implies a possible climate regime shift (e.g., other inter-decadal oscillations take place), suggesting a future updating of dipole predictors may be necessary.

The hindcasting results at two selected lead times (four and eight months) for the JJA precipitation show comparably intermittent skills. For UACF, the better correspondence between the observations and the forecasting envelopes can be seen in early 1980s and late 1990s to early 2000s [see **Figures 4.4(a)** and **4.4(b)**]. The forecasting skills seem to

be better for LACF [see **Figures 4.8(a)** and **4.8(b)**] due to the lower MAE and higher Re values. Furthermore, as mentioned in the last section, if several poor predictions of strong tropical storms in certain years (e.g., 1994 and 2005) can be neglected, the overall forecasting skill for both UACF and LACF strengthens considerably. Regarding the identified dipoles, some differences between the active oceanic regions for UACF and those for LACF are revealed again even if the characteristics of forecasting skills and time series are quite similar. At either lead time, while the North Pacific Ocean is likely to be active for both UACF and LACF, the Atlantic components are much more emphatic for LACF [see **Figures 4.6(a), 4.6(b), 4.10(a),** and **4.10(b)**]. In fact, the Atlantic dipoles shown in **Figure 4.10(a)** bear some similarities with the two recently found Atlantic modes (although these modes are identified in the concurrent period) that may influence the southeastern summer precipitation (Wang et al., 2010). The negative pole in the vicinity of the New England coast refers to the second Atlantic mode (**Fig. 6e** in the Wang et al. paper), and the positive pole in the subtropical Atlantic Ocean resembles the first Atlantic mode (**Fig. 6c** in the Wang et al. paper). This finding implies that the Atlantic modes have not only the concurrent correlation with the southeastern precipitation but also the precursory information. Actually, according to the dipole illustrations, the optimal forecasting of the JJA precipitation can only be achieved by incorporating the simultaneous interaction between the Pacific and Atlantic Oceans in the forecasting model.

Lastly, the hindcasting results at two selected lead times for the SON precipitation are plotted in **Figures 4.4(c)** and **4.4(d)** for UACF and **Figures 4.8(c)** and **4.8(d)** for LACF.

Likewise, hurricane/tropical storm events that occurred in certain years (e.g., 2002 and 2004) affect the forecasting skill to a certain extent. Notably, for UACF, the top three precipitation records (two tropical storm systems and one intense synoptic/mesoscale system) took place in the last ten years. Although the SSTA dipoles during these years generated some responses to the anomalously high precipitation, their magnitude appeared to fall short. Because of the minor impact of the aforementioned extremes on the SON precipitation in LACF, the overall magnitude is less, and the forecasting skills are slightly better. The corresponding dipoles for UACF and LACF [**Figures 4.6(c)**, **4.6(d)**, **4.10(c)**, and **4.10(d)**] only exhibit little dissimilarity for this season. For LACF, the major negative pole in the North Atlantic Ocean (at the 3-month lead time) shifts slightly eastward, and the major positive pole in the Pacific Ocean (at the 4-month lead time) is more dispersive and closer to the West Coast of North America in comparison with those for UACF. In addition, most of the dipole components in the Pacific Ocean seem to be still related to ENSO, which indicates a prolonged impact of ENSO on the southeastern US climate. For example, an ENSO signal detected in mid-spring to summer may determine the climatic condition in fall, and (if) the same signal detected in fall may sequentially determine the climatic condition in winter. Nevertheless, as addressed before, the conjunctive use of both the Pacific and Atlantic oceans can ensure optimal forecasting skills, as identified by the dipole method.

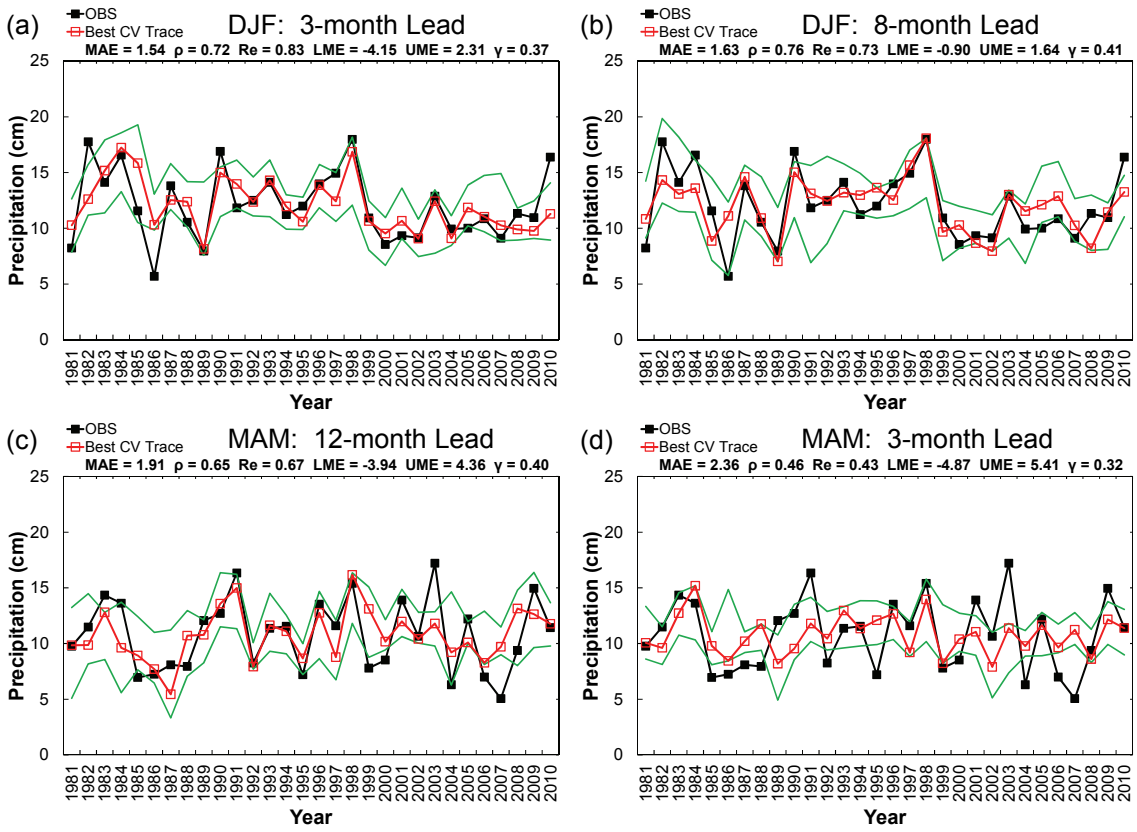



**Figure 4.3:** Optimal hindcasting results at two selected lead times for DJF and MAM precipitation in UACF. The best CV trace is shown in red, and the green envelope is constructed using the maximum and minimum hindcasting values.

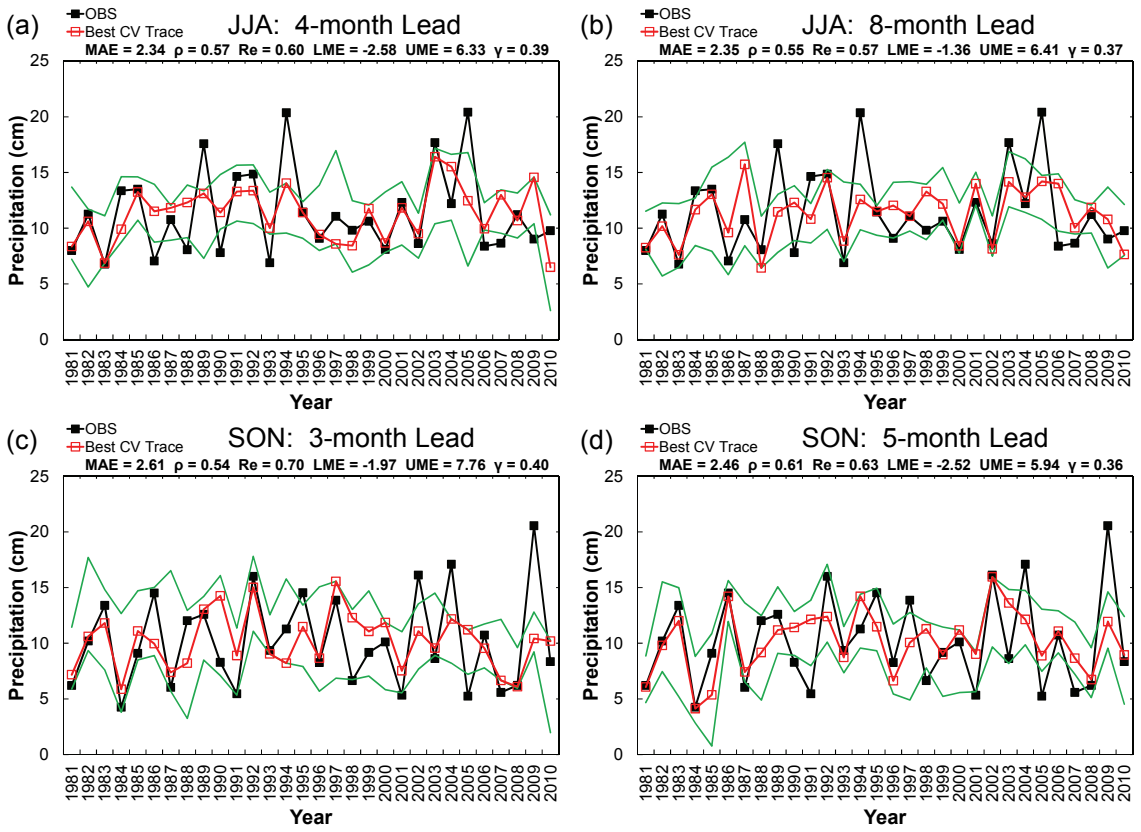



**Figure 4.4:** As in **Figure 4.3**, but for JJA and SON precipitation.

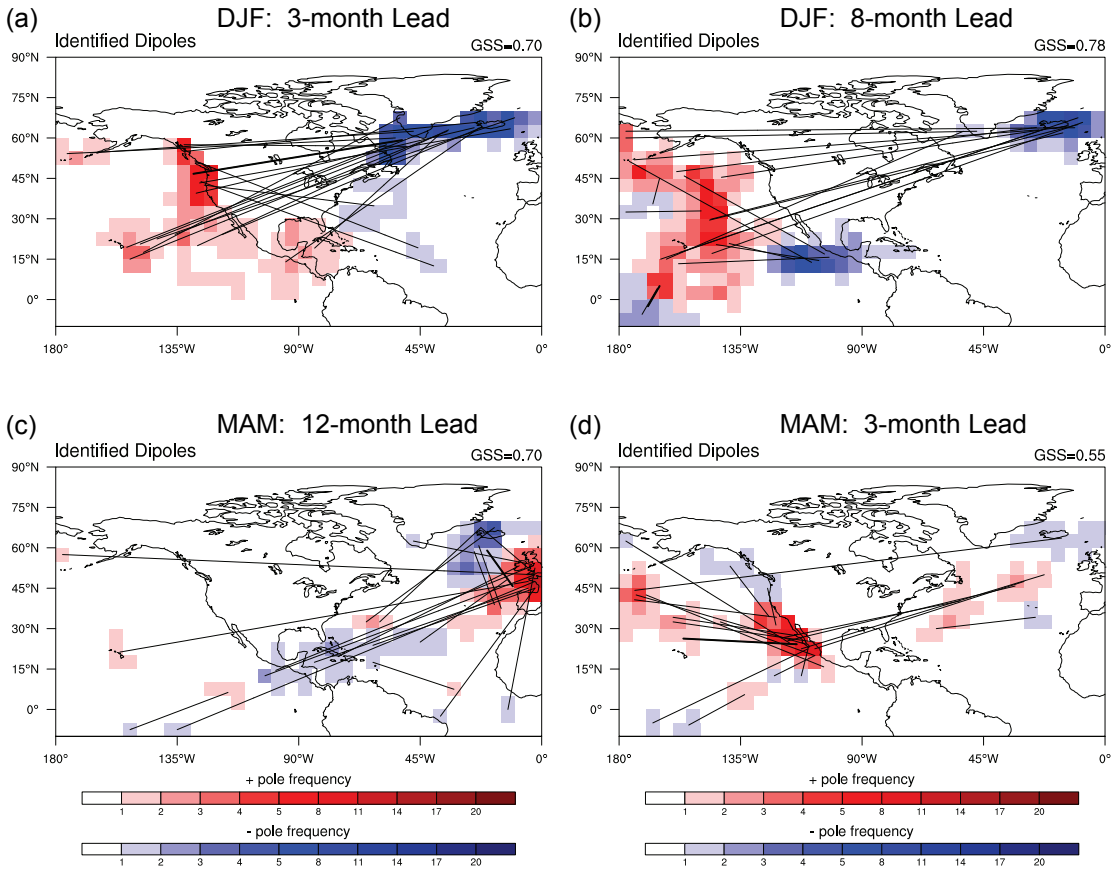



**Figure 4.5:** Corresponding dipoles used to generate the hindcasting results shown in **Figure 4.3**. Thicker and thinner connective lines indicate the orientation of the best and remaining dipoles, respectively. The best GSS among all dipoles obtained during the predictor screening process is denoted at the upper-right corner of each figure.

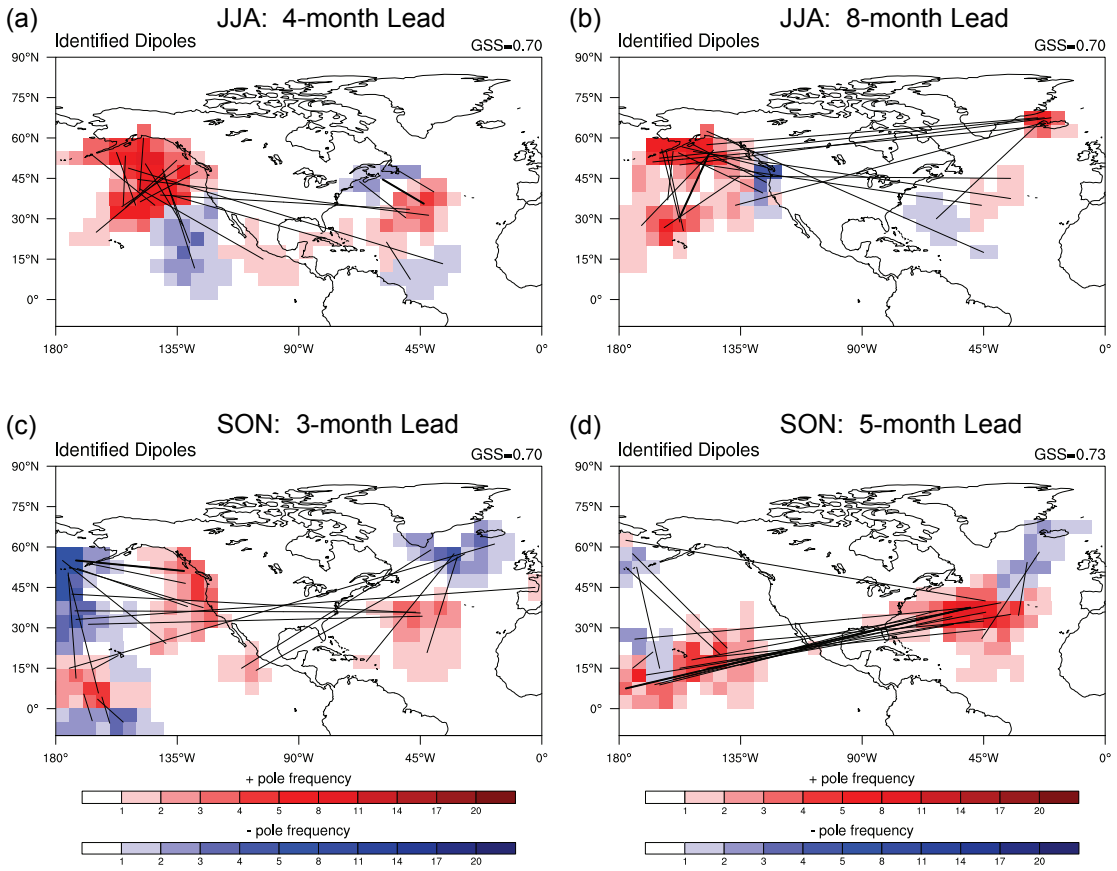



**Figure 4.6:** As in **Figure 4.5**, but for the hindcasting results shown in **Figure 4.4**.

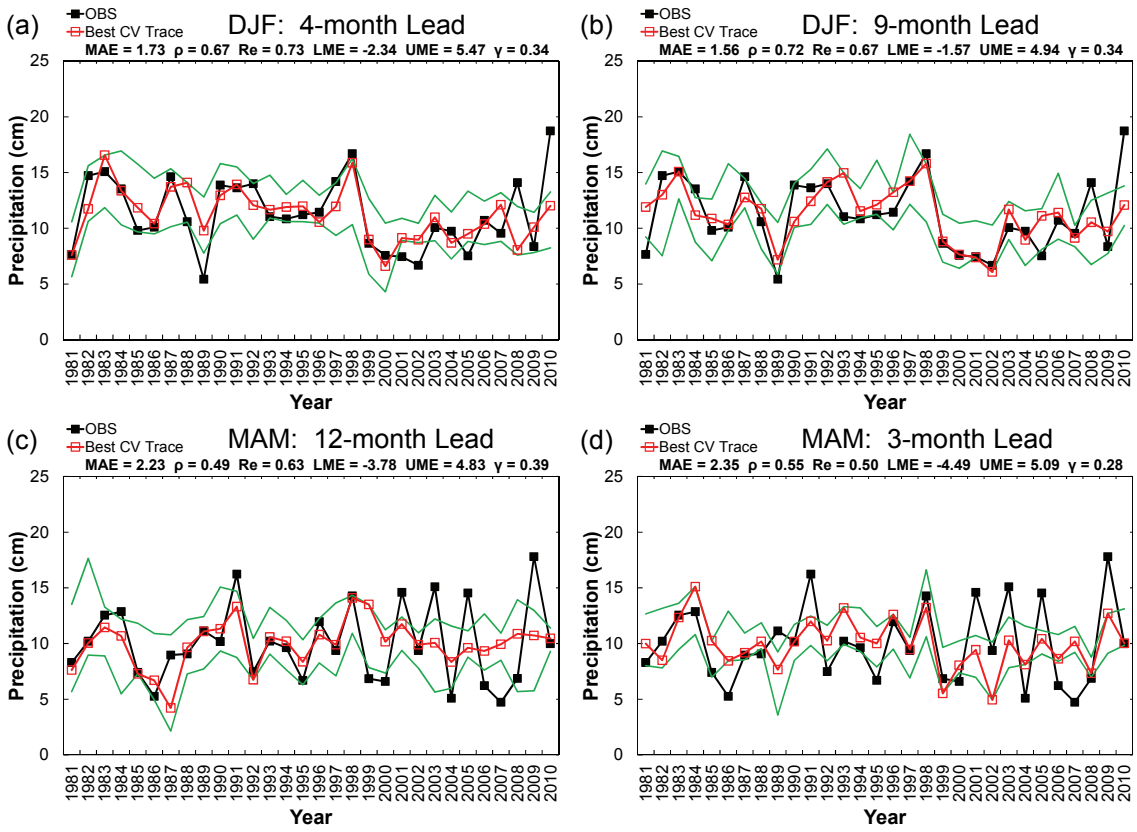



**Figure 4.7:** As in **Figure 4.3**, but for DJF and MAM precipitation in LACF.

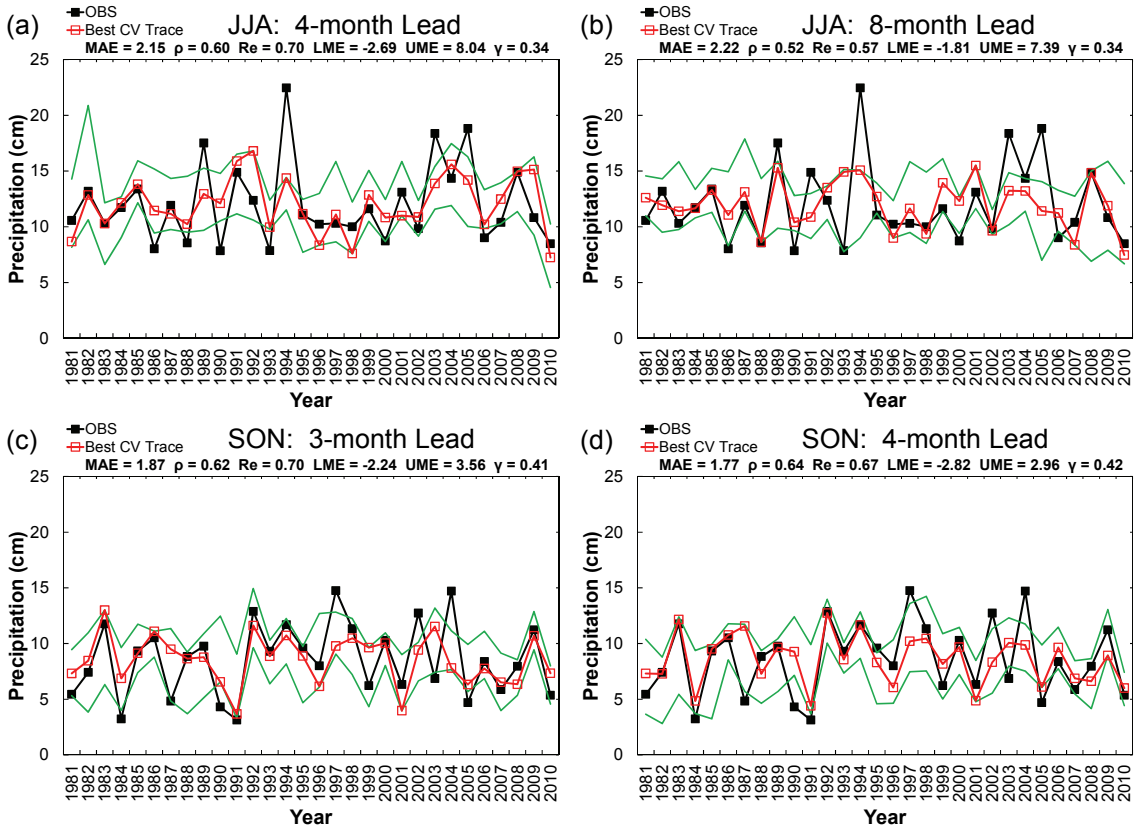



**Figure 4.8:** As in **Figure 4.4**, but for JJA and SON precipitation in LACF.

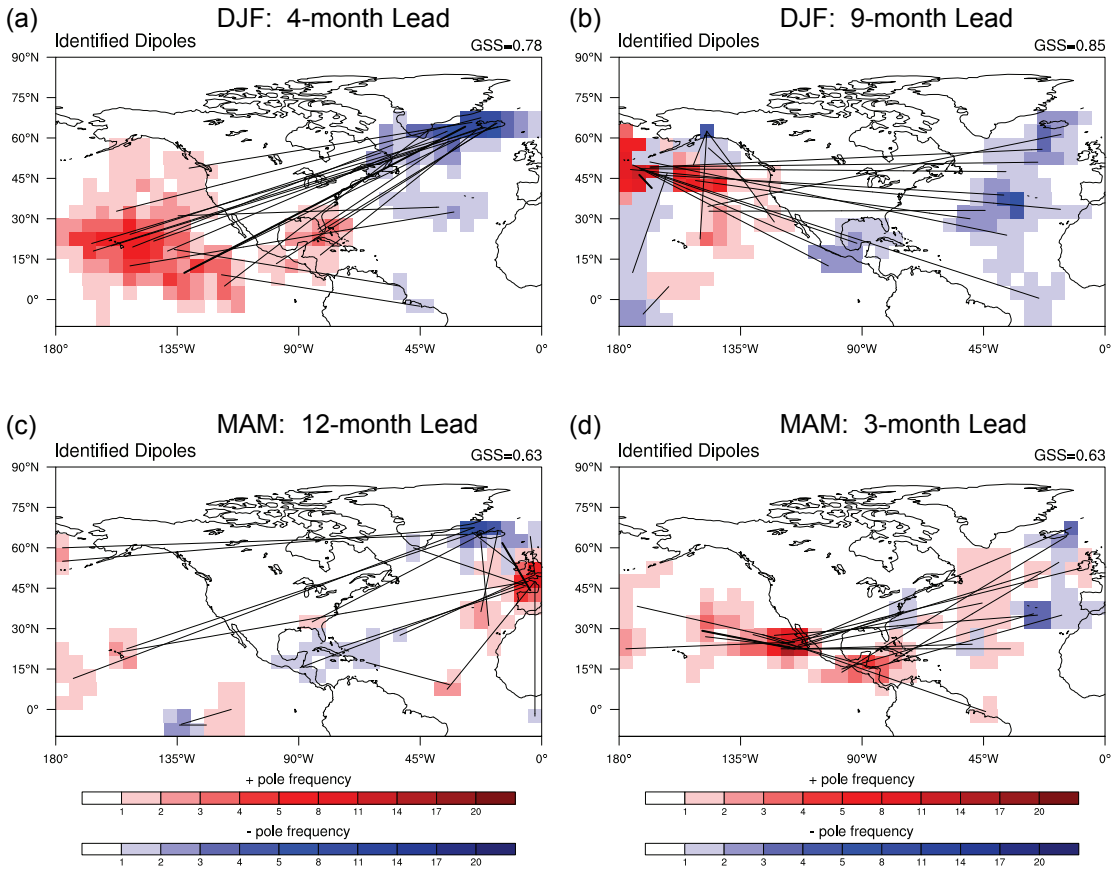



**Figure 4.9:** As in **Figure 4.5**, but for LACF.

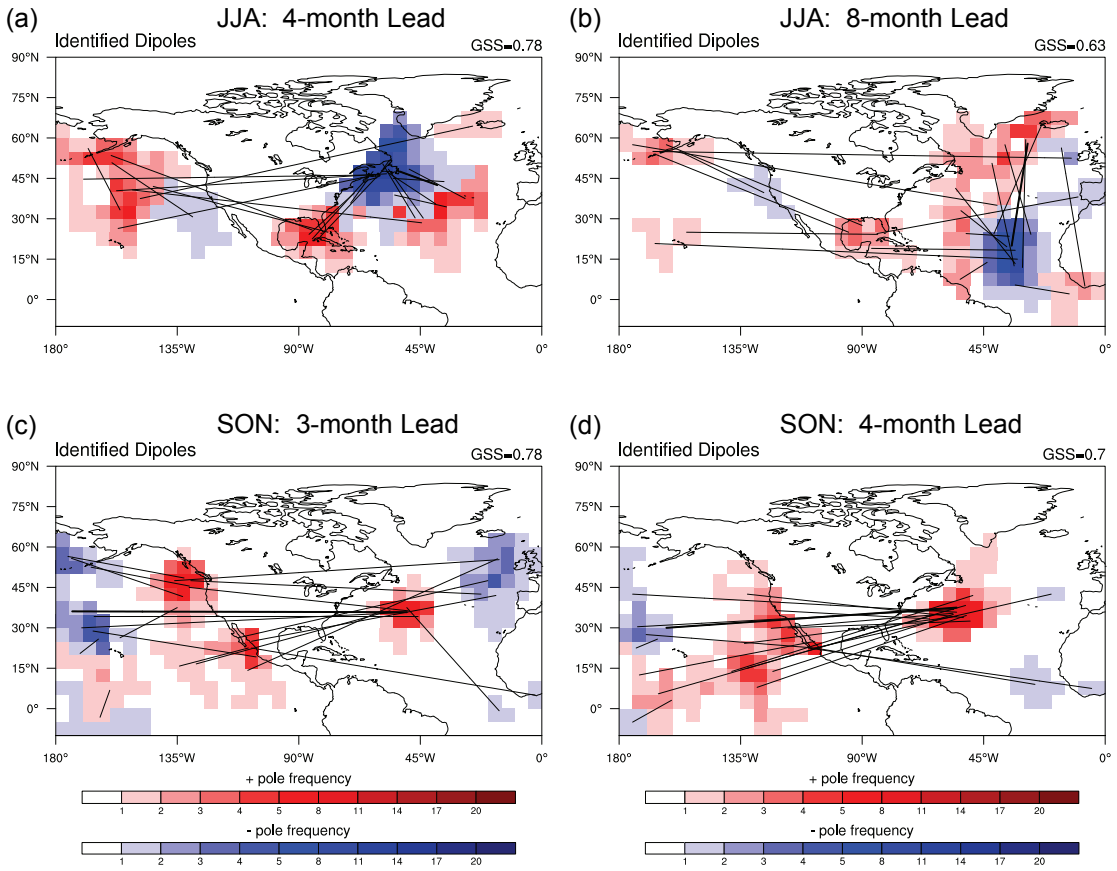



**Figure 4.10:** As in **Figure 4.6**, but for LACF.

### 4.3 Intercomparison of Forecasting Schemes

To verify whether the dipole method is robust, the hindcasting skills (with emphasis on $MAE_{CV}$ and $\rho_{CV}$) presented in the previous section are compared with the skills generated by other statistical methods. The first method adopted is canonical correlation analysis (CCA), generally acknowledged as the most effective statistical method operationally used in the CPC. Because CCA is guaranteed to explain the maximum variance in the predictand field through finding the optimum linear combination of the predictor fields, it has been extensively applied to many climate forecasting studies (e.g., Barnett and Preisendorfer, 1987; Barnston and Ropelewski, 1992; Barnston, 1994; Barnston and He, 1996; Yu et al., 1997; Johansson et al., 1998; Ward, 1998; Landman and Mason, 1999; Wilks, 2008). Thus, results generated by CCA pass for possessing benchmark performance among the intercomparison processes (More computational details regarding CCA are given in the appendix).

For the purpose of examining the influence of ENSO on precipitation in the Southeast, the second method adopted is a simple linear regression scheme based on an ENSO index. The ENSO index used here is NINO 3.4, which is the SSTA series retrieved from a bounded area (5°S–5°N and 120°W–170°W) and available on the CPC FTP server (ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/sstoi.indices). The last method adopted is climatological (CLM) forecasting; that is, a constant forecast based on the mean value of long-term precipitation in each season. Skills generated by CLM forecasting can help to justify whether the above methods present true skills. Moreover, to evaluate the best CV

hindcasting skills, these methods also sift through various lead time and undergo a similar CV procedure as applied to the dipole method.

After applying all these methods to the same case study, the hindcasting skills for all methods (only the best results based on the lowest MAEs among all lead times are shown) are summarized in **Table 4.5**. Generally, in all seasons, none of the common statistical methods can outperform the dipole method in either $MAE_{CV}$ or $\rho_{CV}$. Although CCA shows some skills and similarity at the best lead times among all seasons, these skills are not comparable to those generated by the dipole method. Regarding ENSO regression, the only season when it exhibits skills is DJF. In other seasons, ENSO regression provides no skills since almost all the derived MAEs are even higher than those derived from CLM forecasting, and the correlation coefficients are all nearly zero. As a result, the corresponding best lead times basically reveal no substantial meaning.

The reasons that the dipole method compares favorably with both CCA as well as ENSO regression are as follows. Because the nature of CCA is to utilize complete spatial variables (i.e., SSTA in each grid) to generate the optimal linear combination of predictors and predictands, the risk that CCA is overfit is high. In contrast, ENSO regression uses only one fixed spatial predictor, so the ENSO index sometimes fails to show any significant relationships with predictands of interest. Overall, the dipole method, which identifies the most significant oceanic areas as predictors and undergoes a strict CV procedure to prevent overfitting, appears to be a more robust statistical tool for climate forecasting.

**Table 4.5:** Comparisons between the hindcasting skill ( $MAE_{CV}$ , $\rho_{CV}$ ) of the dipole (DP) method and the skills of other operational or common methods.

| Season | Region | Method | | | |
|---|---|---|---|---|---|
| | | DP* | CCA* | ENSO* | CLM† |
| DJF | UACF | 1.54, 0.73 (3) | 2.35, 0.44 (6) | 2.35, 0.34 (0) | 2.59, *NA* |
| | LACF | 1.56, 0.72 (9) | 2.28, 0.38 (2) | 2.26, 0.49 (0) | 2.73, *NA* |
| MAM | UACF | 1.91, 0.65 (12) | 2.63, 0.33 (12) | 2.76, -0.06 (0) | 2.83, *NA* |
| | LACF | 2.23, 0.49 (12) | 3.00, 0.22 (12) | 2.91, -0.12 (2) | 2.89, *NA* |
| JJA | UACF | 2.32, 0.56 (12) | 2.55, 0.43 (12) | 3.10, -0.13 (0) | 3.03, *NA* |
| | LACF | 2.04, 0.56 (3) | 2.48, 0.42 (12) | 2.87, -0.08 (0) | 2.83, *NA* |
| SON | UACF | 2.46, 0.61 (5) | 3.41, 0.25 (1) | 3.43, 0.09 (0) | 3.50, *NA* |
| | LACF | 1.77, 0.64 (4) | 2.28, 0.40 (1) | 2.78, 0.03 (0) | 2.78, *NA* |

*: Numbers in parentheses represent the best lead times (in month) corresponding to each season

†: $\rho$ is not available ($NA$) for CLM

## 4.4 Diagnostic Analysis of Large-Scale Variables

To clarify the possible physical relationships between the identified dipoles and large-scale circulation, additional atmospheric variables are examined. More specifically, geopotential height and vector wind data at a 500mb level retrieved from the NCEP/NCAR Reanalysis dataset (Kalnay et al., 1996) are selected to generate composite maps. Wet and dry composite years in accordance with the four seasons and two regions are distinguished first (**Table 4.6**). The wet or dry years are defined as years when seasonal precipitation falls in the high or low tercile of distribution. Notably, some of the wet and dry events tend to persist more than one season. For example, during the period 1983 to 1984, both UACF and LACF continued to be in wet conditions, but during the period 2006 to 2007, the ACF basin as well as the entire southeastern US experienced a severe long lasting drought.

**Table 4.6:** Wet and dry years (corresponding to different seasons and regions) used in composite analysis.

| Season | Region | Composite Years | |
|---|---|---|---|
| | | Wet | Dry |
| DJF | UACF | 1982, 1983, 1984, 1987, 1990, 1993, 1996, 1997, 1998, 2010 | 1981, 1986, 1988, 1989, 2000, 2001, 2002, 2004, 2005, 2007 |
| | LACF | 1982, 1983, 1987, 1990, 1991, 1992, 1997, 1998, 2008, 2010 | 1981, 1989, 1999, 2000, 2001, 2002, 2004, 2005, 2007, 2009 |
| MAM | UACF | 1983, 1984, 1990, 1991, 1996, 1998, 2001, 2003, 2005, 2009 | 1985, 1986, 1987, 1988, 1992, 1995, 1999, 2004, 2006, 2007 |
| | LACF | 1983, 1984, 1989, 1991, 1996, 1998, 2001, 2003, 2005, 2009 | 1985, 1986, 1992, 1995, 1999, 2000, 2004, 2006, 2007, 2008 |
| JJA | UACF | 1984, 1985, 1989, 1991, 1992, 1994, 2001, 2003, 2004, 2005 | 1981, 1983, 1986, 1988, 1990, 1993, 2000, 2002, 2006, 2007 |
| | LACF | 1982, 1985, 1989, 1991, 1994, 2001, 2003, 2004, 2005, 2008 | 1986, 1988, 1990, 1993, 1996, 1998, 2000, 2002, 2006, 2010 |
| SON | UACF | 1983, 1986, 1988, 1989, 1992, 1995, 1997, 2002, 2004, 2009 | 1981, 1984, 1987, 1991, 1996, 1998, 2001, 2005, 2007, 2008 |
| | LACF | 1983, 1986, 1992, 1994, 1997, 1998, 2000, 2002, 2004, 2009 | 1981, 1984, 1987, 1990, 1991, 1999, 2001, 2005, 2007, 2010 |

### *4.4.1 Upper-Level Fields in Concurrent Seasons*

DJF composite maps for both UACF and LACF (**Figure 4.11**) display two negative 500mb geopotential height anomalies that occupy the northern Pacific and Atlantic basins and another positive anomaly centered on eastern Canada during the wet years. The locations of these centers seem to have some connections with the identified dipoles for UACF at a 3-month lead time. According to these patterns, the southeastern US precipitation is influenced by an anomalous trough that extends from the northern Pacific basin and induces southerly or southwesterly flow anomalies that provide moisture

supply from the eastern Pacific and the Gulf of Mexico. In contrast, DJF composite maps for the dry years present a nearly out-of-phase structure, that is, positive height anomalies prevail over the northern Pacific basin, North America, and the northern Atlantic basin. Thus, the patterns of wet-minus-dry years [**Figures 4.11(e)** and **4.11(f)**], which emphasize anomalous contrasts at the same geographical location, are similar to the patterns of the wet years.

Patterns in MAM (**Figure 4.12**) present some differences compared to those in DJF. In the wet years, zonal positive height anomalies traverse high latitude regions from the Gulf of Alaska, eastern Canada, and Iceland. In the dry years, a clear opposite phase of height anomalies can be found in the vicinity of Alaska. The continental height anomalies and associated air flows shown in wet-minus-dry composite maps [**Figures 4.12(e)** and **4.12(f)**] still contribute to some variability in southeastern US precipitation; however, this relationship and associated patterns in MAM compared to those in DJF are weakened. Moreover, because the identified dipoles for this season are more diverse, any correlation between oceanic dipoles and atmospheric variables can hardly be seen. All the results indicate that variability in the MAM precipitation in the southeastern US barely responds to either oceanic forcing or synoptic-scale systems.

JJA composite maps display more disorganized patterns in weak anomalies except a pronounced anomalous Arctic low in the UACF wet years [**Figure 4.13(a)**]. From the composite maps of wet-minus-dry years [**Figures 4.13(e)** and **4.13(f)**], multiple anomalous low centers similar (but much weaker) to patterns in DJF that influence wet

and dry conditions for both UACF and LACF can still be seen. Apart from extratropical weather systems, other tropical disturbances (e.g., hurricanes) are shown to influence summer storminess in the southeastern US. It is argued that the genesis and track variability of hurricanes should be partly attributed to both the identified dipoles and the upper-level patterns in such sensitive locations as the New England coast, the subtropical Atlantic Ocean, and the northern Atlantic. However, the resemblance between the geographic locations of these dipoles and patterns is low.

Patterns in SON (**Figure 4.14**) compared to those in JJA demonstrate stronger extratropical influence. At North America, the condition of continental height anomalies still plays a role in moisture transport. Over the North Pacific, a meridional geopotential height dipole centered on the Arctic Ocean and the Gulf of Alaska seems to be important as well. A noticeable difference between the patterns of UACF and those of LACF during the wet SON years is the clear opposite phase of height anomalies in eastern Canada. Whereas the patterns of UACF indicate that negative height anomalies reside in the eastern/northeastern Canada region (and stretch down to the southeastern US), the patterns of LACF indicate that positive height anomalies occupy the same region. As a result, what influences the LACF climate instead is an anomalous trough originated at the West Coast. Therefore, southerly/southeasterly flow anomalies near the Gulf of Mexico depict some zonal displacement between these two cases. From the Atlantic side, similar to the JJA patterns, the persistent height anomalies in the subtropical Atlantic Ocean seem to partially coincide with the identified dipoles and may be also related to hurricane activities during SON.

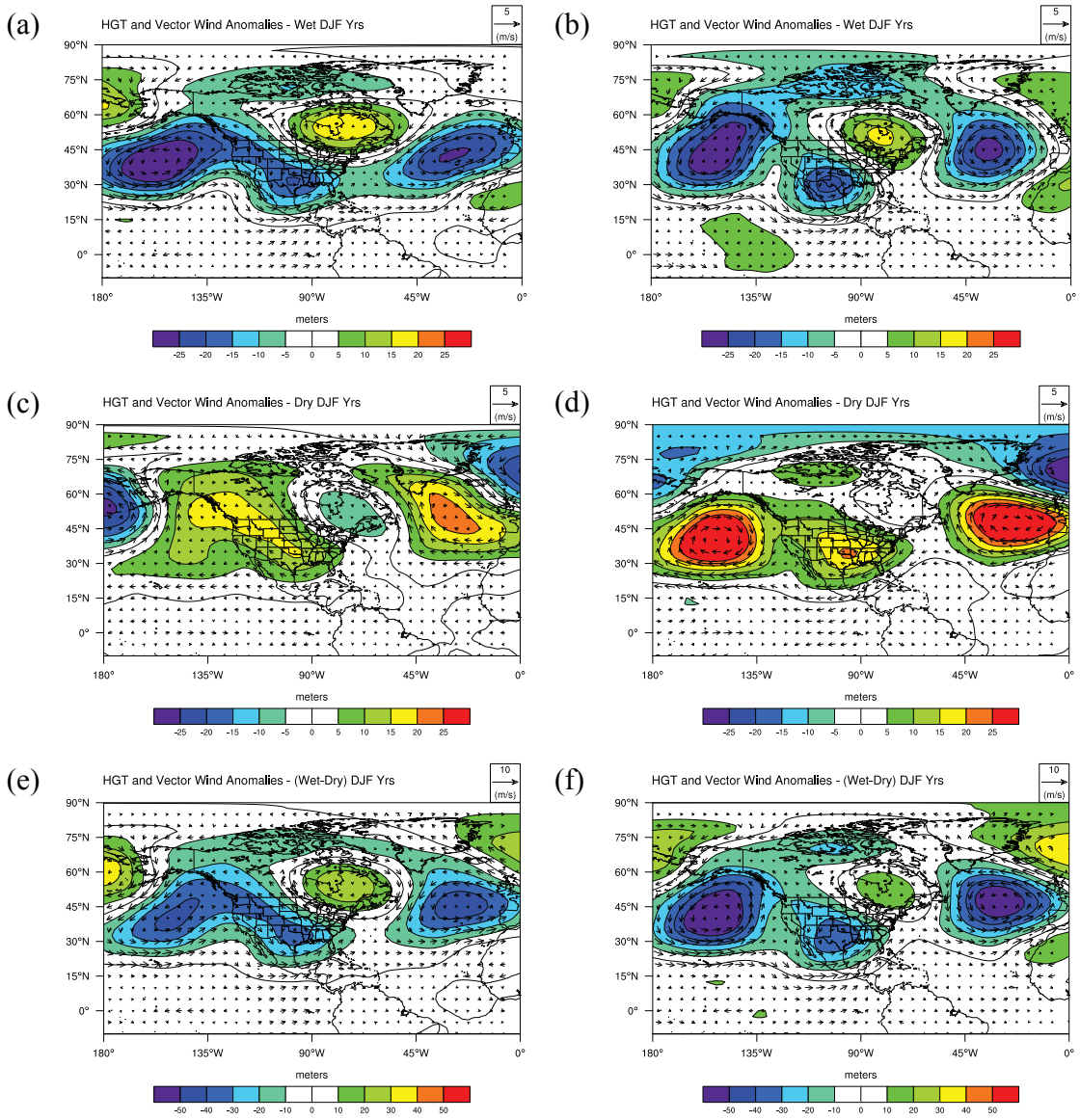



**Figure 4.11:** DJF composite maps of NCEP/NCAR Reanalysis data for wet [(a) and (b)], dry [(c) and (d)], and wet-minus-dry years [(e) and (f)]. Left and right panels are based on UACF and LACF composite years, respectively. Shading with overlying contours are 500mb geopotential height anomalies, and arrows are associated vector wind anomalies.

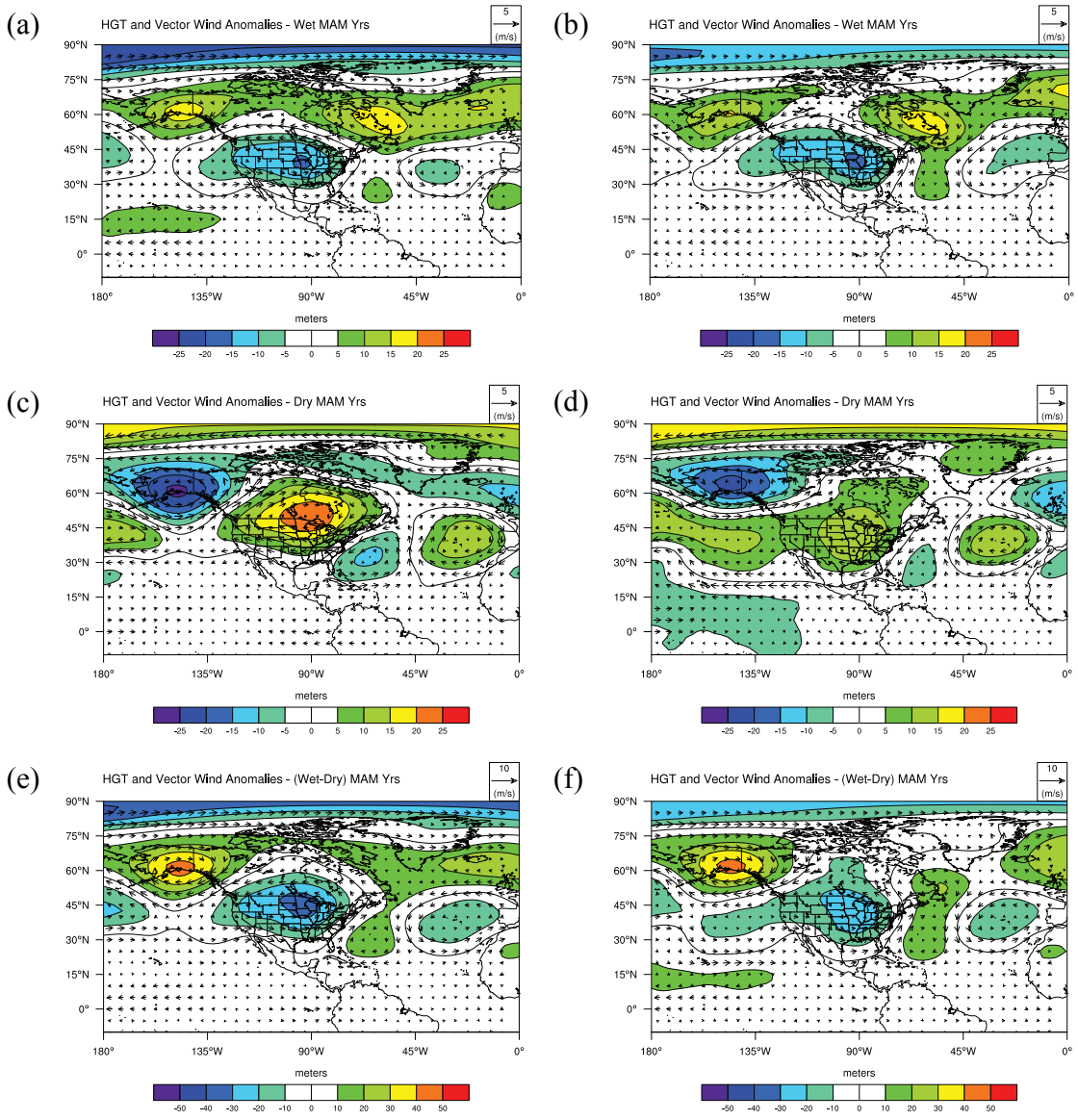



**Figure 4.12:** As in **Figure 4.11**, but for MAM.

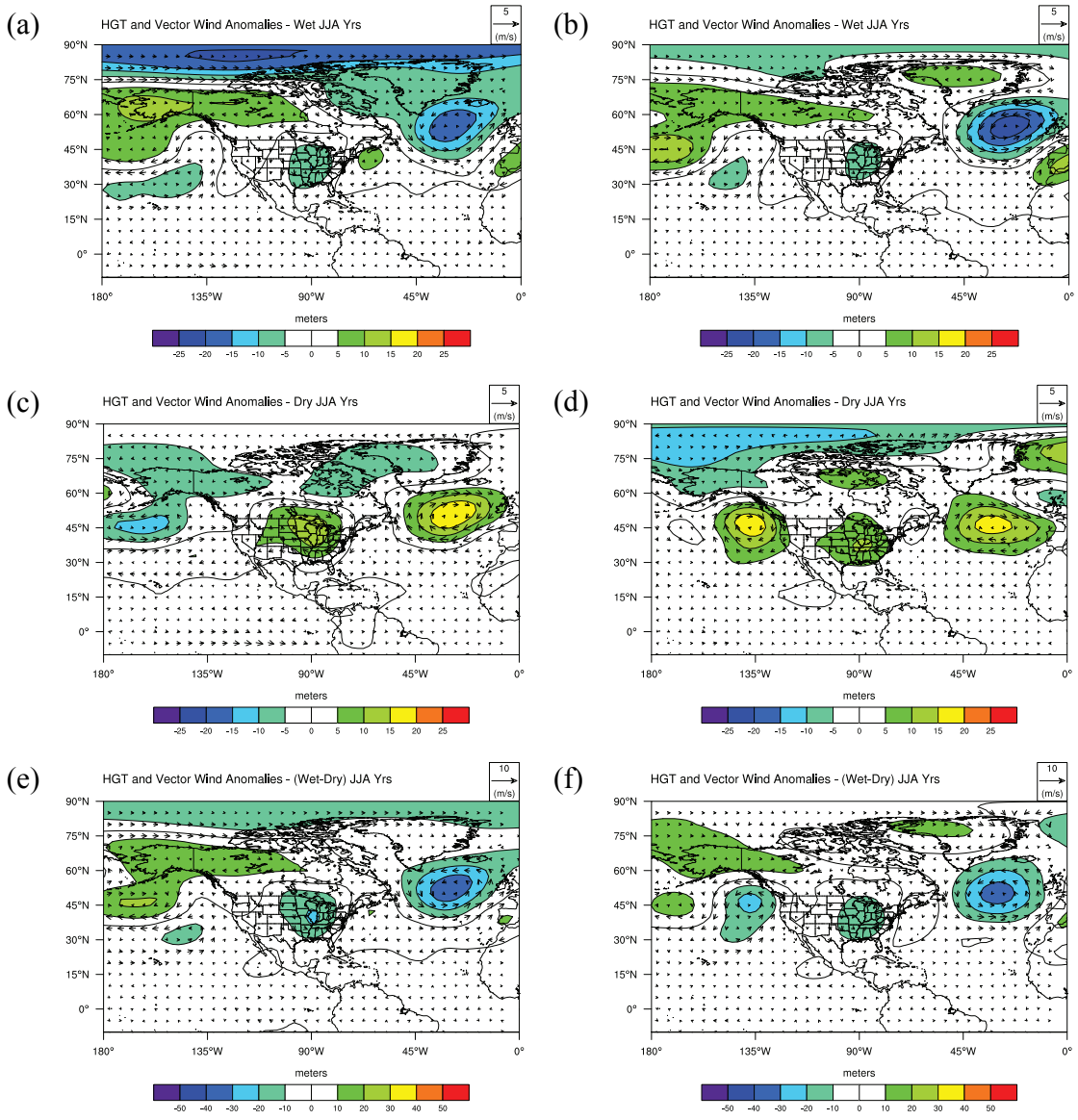



**Figure 4.13:** As in **Figure 4.11**, but for JJA.

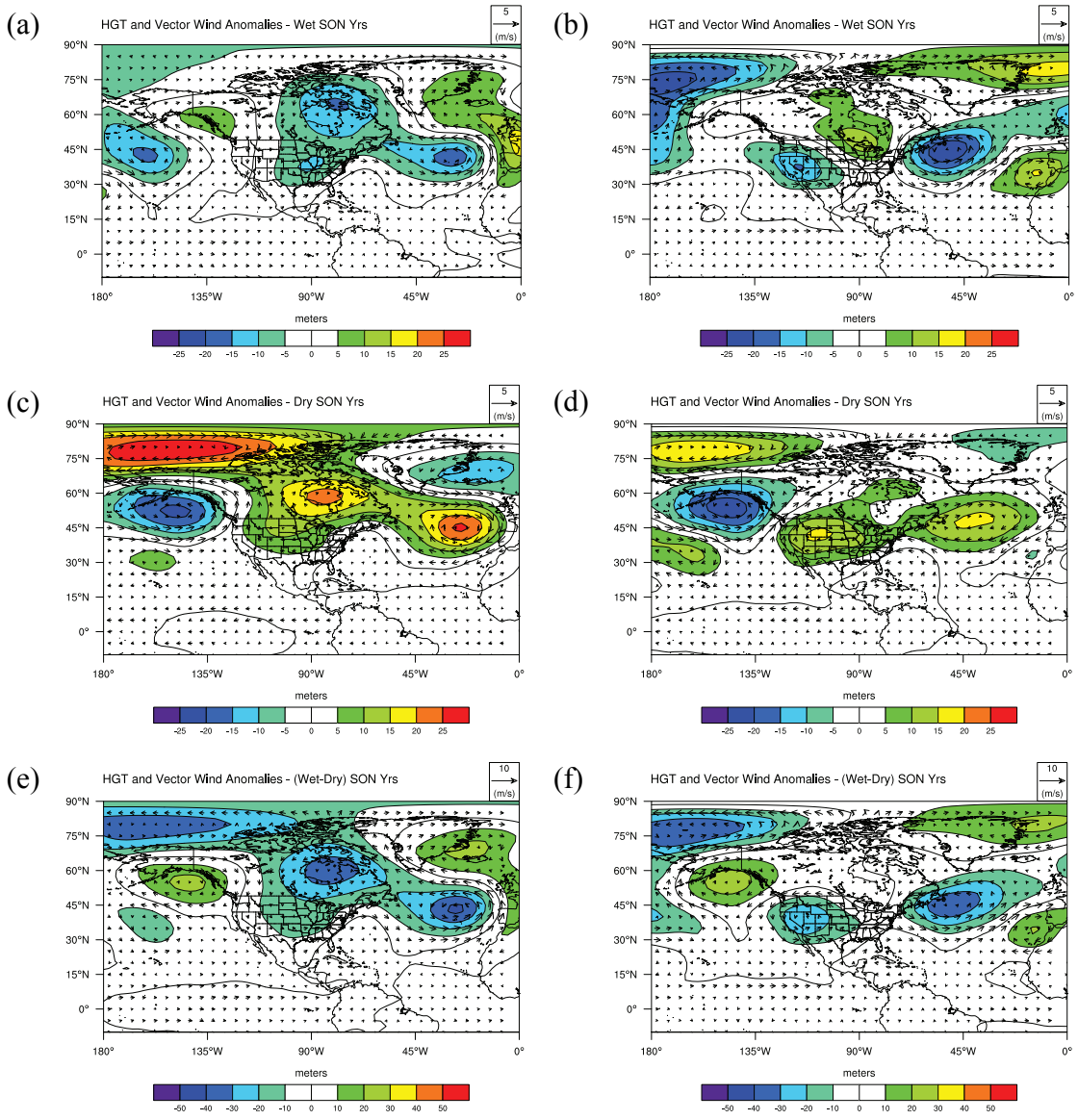



**Figure 4.14:** As in **Figure 4.11**, but for SON.

*4.4.2 Surface Fields in Corresponding Lead Times*

To examine whether the identified SSTA dipoles at the 3- and 4-month lead times are physically meaningful, SSTA composite maps are generated. Based on the wet and dry DJF years for UACF listed in **Table 4.6**, **Figure 4.15** presents the DJF composite maps (with a 3-month lead time) of the Kaplan SSTA along with the NCEP/NCAR Reanalysis SLP and surface vector wind anomalies. In the wet DJF years, an El Niño signal and a prominent low SLP anomaly at the vicinity of the Gulf of Alaska [**Figures 4.15(a)** and **4.15(b)**] can be seen in the preceding season. In the dry DJF years, while the SLP anomaly in the same location displays a slightly opposite phase, the SSTA in the eastern tropical Pacific indicates a normal to slightly warm condition. However, a cold SSTA appears at the Northwest Coast of North America, resulting in a clear contrasting anomaly in the West Coast as shown in the wet-minus-dry composite map [**Figures 4.15(e)]**. The contrasting anomaly in the West Coast along with the other contrasting SSTA pattern from the Labrador Sea to the coast of Iceland were indeed identified by the proposed algorithm as the most predictive dipoles [see **Figures 4.5(a)]** in lieu of the ENSO pattern.

For UACF, the algorithm did not identify or include the key ENSO region as the most predictive dipoles for the following reasons: 1) Because of the characteristic of the GSS, the algorithm tends to find SSTA regions that show the highest contrast between the "dipole values" in the wet years and those in the dry years; 2) Even though the composite maps indicate that most of the wet (dry) DJF years coincide with the El Niño (La Niña) events, several years with neutral to opposite conditions can be observed. This may

drastically decrease the forecasting skill if the key ENSO region is adopted as the predictor; and 3) Because the number of final dipole selection is specified as $\xi = 20$, it may not be enough to keep dipoles potentially located in the tropical regions (with comparatively low MAEs).

Likewise, the similar DJF composite maps of the surface field data for LACF (with a 4-month lead time) is plotted in **Figure 4.16**. In the wet DJF years, the El Niño signal can still be seen, but the low pressure anomaly in the northeastern Pacific, unlike the strong anomaly observed for UACF, seem to be less pronounced. In the dry DJF years, the Pacific Ocean seems to be close to neutral and even slightly colder than normal. Thus, although the warm contrasting anomaly still appears at the West Coast as shown in **Figure 4.16(e)**, the algorithm identified a vast Pacific region as the most predictive positive pole for LACF. Moreover, the cold contrasting anomaly appears to be more active at the vicinity of Iceland, which explains why the identified negative pole for LACF (compared with that for UACF) shows some eastward displacement. Overall, the composite map of the SSTA can approximately pinpoint predictor regions but not predictor structures or orientations, which manifests the utility of the dipole identification algorithm.

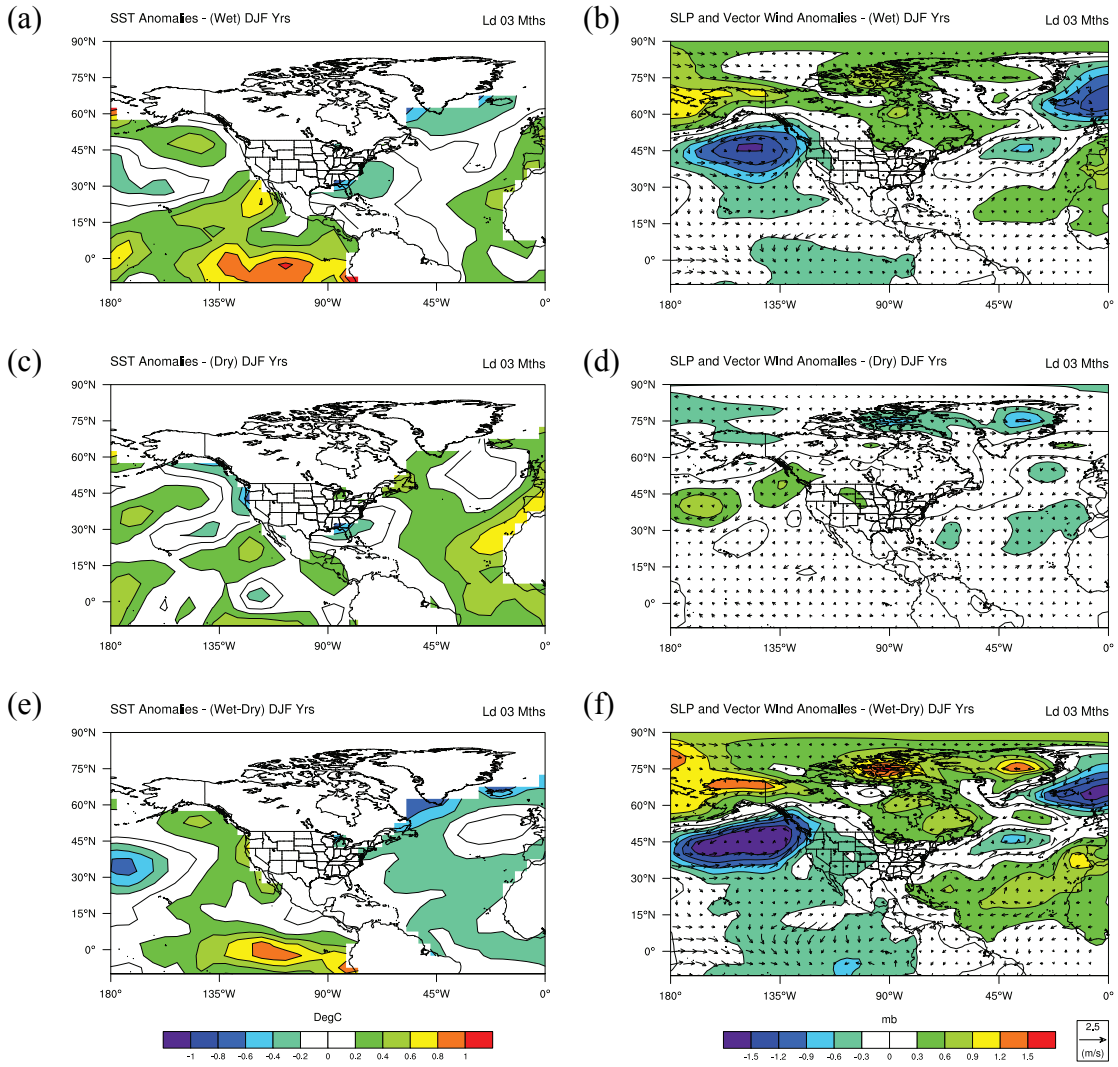



**Figure 4.15:** DJF composite maps (with a 3-month lead time) of SSTA (left panel) and SLP and surface wind anomalies (right panel) for wet [(a) and (b)], dry [(c) and (d)], and wet-minus-dry years [(e) and (f)] for UACF.

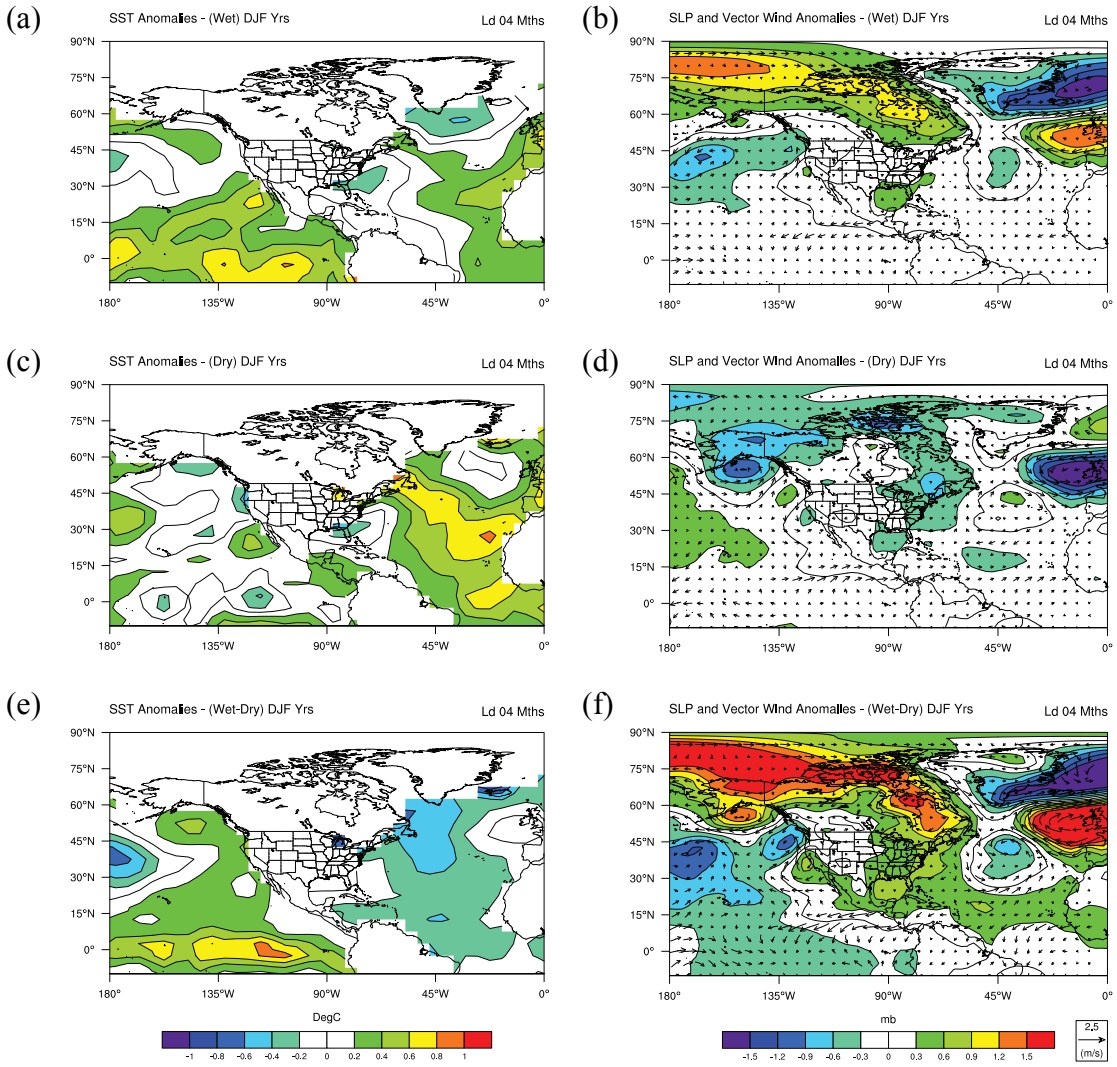



**Figure 4.16:** As in **Figure 4.15**, but for LACF (with a 4-month lead time).

## 4.5 Demonstration of 2012 Forecasts

The last section of this chapter demonstrates an operational forecasting framework for the year-round seasonal precipitation in the ACF basin. This forecasting framework uses the newest available SSTA data (January 2012 when this work is finalized) from which the best dipole predictors at certain lead times (according to the highest hindcasting Re values) are selected to generate the 2012 forecasts. Because the SSTA data at corresponding lead times may not be always available, it is not guaranteed to select the dipole predictors at the optimum lead times (based on the hindcasting experiment described in Section 4.2). However, a progressive routine that can instantly update forecasts whenever new SSTA data become available has been developed.

In light of this framework, the forecasting results of 2012 seasonal precipitation are generated. The main statistics of the forecasts along with the long-term climatology (derived from 1981 to 2010) are summarized in **Table 4.7**. To better visualize the statistics, the distribution of the forecasts is also presented in histograms (**Figures 4.17** and **4.18**) and boxplots (**Figure 4.19**). The results show that the year 2012 may start out dry, based on the SSTA dipole predictors in SON and ASO for UACF and LACF, respectively. These forecasts are provided with more than 70% reliability referring to the hindcasting experiment and have been deemed as the successful forecasts compared with the newly observed precipitation.

The spring forecasts, based on the SSTA dipole predictors in previous MAM, indicate a possible upturn in precipitation with about 60–70% reliability. More than half of the

forecasting traces suggest a wet scenario ($\geqq$ the 67th percentile), recharging water supplies for major reservoirs and rivers in the ACF basin.

Nevertheless, the above-normal forecasts do not seem to persist across the seasons. The summer forecasts, based on the SSTA dipole predictors in previous SON and OND for UACF and LACF, suggest a dry scenario with about 60% reliability. For UACF, although none of the forecasting traces are lower than the 33rd percentile, 65% are lower than the long-term mean. For LACF, an even drier condition is suggested since 35% of the forecasting traces are lower than the 33rd percentile.

The fall forecasts, based on the SSTA dipole predictors in previous SON and OND for UACF and LACF, show some increased precipitation especially in UACF. However, because a short lead time is more favorable for the forecasting of the fall precipitation according to the hindcasting results, the current forecasting reliability ($\leqq$ 50%) may be too low to convey any definitive information.

Given the fact that the southeast US has been under a drought since last year, 2012 will be likely to be dry from the hydrologic perspective. Thus, ACF water management should be conservative.

**Table 4.7:** Summary of 2012 seasonal precipitation forecasts in the Southeast US.

| Region | Statistics | Seasons | | | |
|---|---|---|---|---|---|
| | | DJF | MAM | JJA* | SON* |
| UACF | Max | 12.41 | 14.42 | 12.19 | 13.32 |
| | Mean | 10.83 | 12.62 | 10.25 | 11.21 |
| | Min | 9.22 | 9.30 | 8.86 | 9.54 |
| | Re | 0.83 | 0.67 | 0.64 | 0.46 |
| | LT Max | 17.98 | 17.21 | 17.68 | 20.55 |
| | LT Mean | 12.04 | 10.79 | 10.66 | 9.71 |
| | LT Min | 5.69 | 5.06 | 6.76 | 4.25 |
| LACF | Max | 13.59 | 13.26 | 13.21 | 11.13 |
| | Mean | 10.35 | 11.70 | 10.95 | 8.39 |
| | Min | 7.19 | 9.42 | 7.48 | 6.72 |
| | Re | 0.73 | 0.63 | 0.61 | 0.50 |
| | LT Max | 18.74 | 17.80 | 18.37 | 14.75 |
| | LT Mean | 11.25 | 9.99 | 11.41 | 8.19 |
| | LT Min | 5.43 | 4.73 | 7.85 | 3.14 |

*: statistics are obtained by excluding major hurricane events

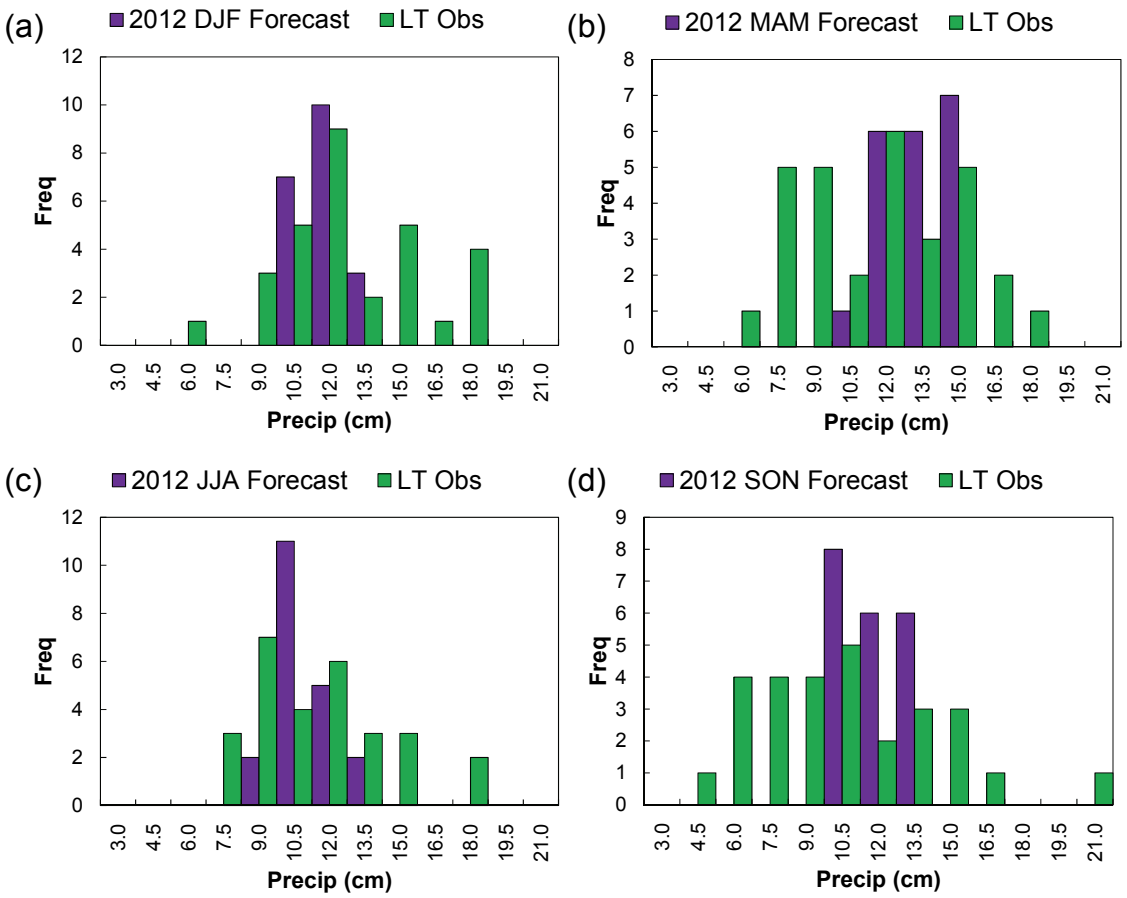



**Figure 4.17:** Histograms of 2012 seasonal precipitation forecasts for UACF.

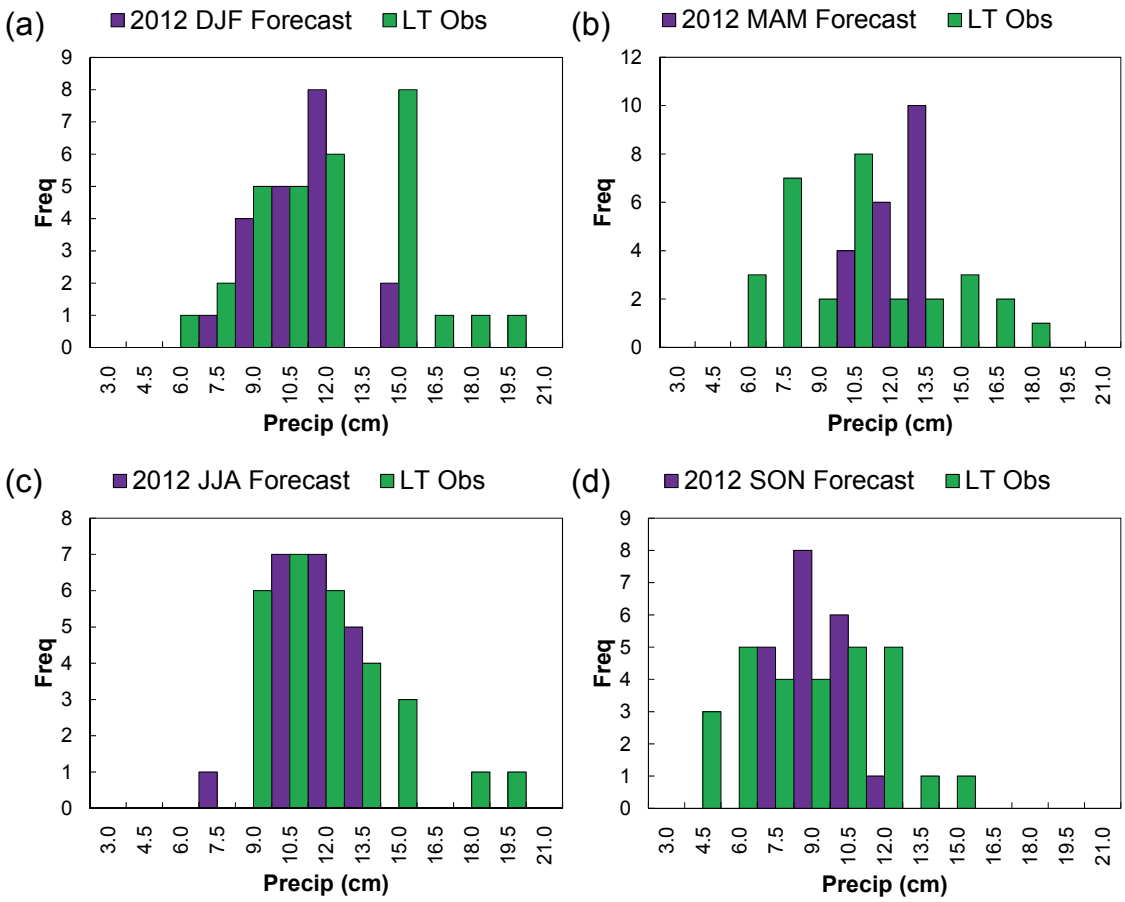



**Figure 4.18:** As in **Figure 4.17**, but for LACF.

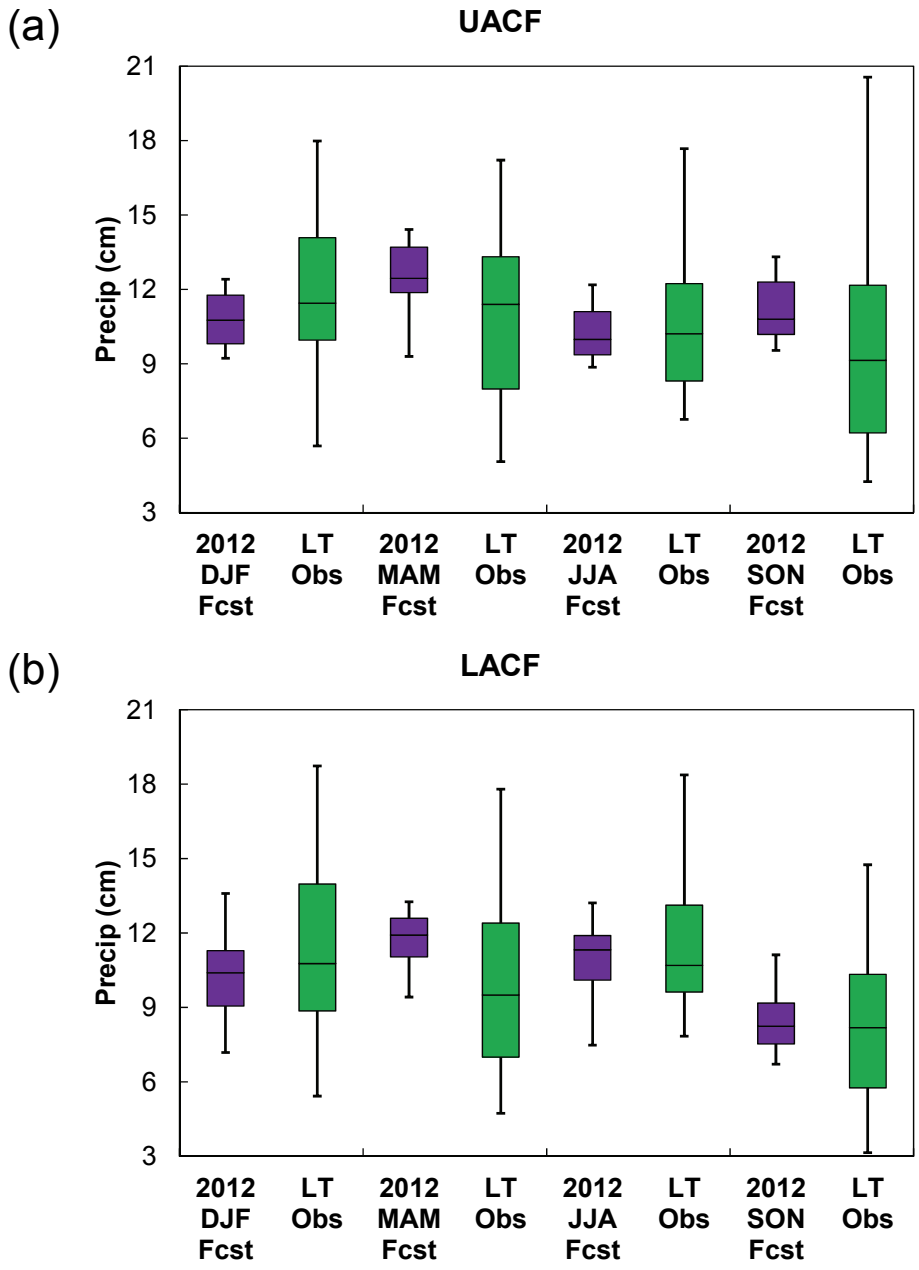



**Figure 4.19:** Boxplots of 2012 seasonal precipitation forecasts for (a) UACF and (b) LACF.

# CHAPTER 5

# CASE STUDY IN EAST AFRICA

The second case study carried out in this thesis pertains to the forecasting of seasonal precipitation in East Africa. This region is highly vulnerable to climate variability, as devastating recent drought has clearly demonstrated. Section 5.1 describes the different East African rainy seasons examined in this chapter. Section 5.2 first presents the general method set-up and then interprets the associated results. Section 5.3 carries out a sensitivity analysis of SSTA domains to inspect if potential dipole predictors and improved forecasting skills can be identified from additional oceanic areas. Section 5.4 compares the results generated by the dipole method with those generated by other statistical methods. Section 5.5 performs climate diagnostics for selected rainy seasons to discuss hypothetical mechanisms and possible relationships between the identified dipoles and the surface and upper-level variables. Section 5.6 demonstrates the latest forecasting results of 2012 seasonal precipitation in East Africa.

## 5.1 East Africa Seasonal Precipitation

The precipitation data used in this study is derived from the CMAP data as fourteen $5 \times 5$ degree grids (**Figure 5.1**) that cover countries from equatorial East Africa (Tanzania, Uganda, and Kenya) to as farther north as countries in the Horn of Africa (Ethiopia and Somalia). Because the East African climate is modulated by the seasonal migration of the Inter-Tropical Convergence Zone (ITCZ), effects of manifold orography (e.g., Ethiopia Highlands, Mountain Kilimanjaro, and the Great Rift Valley), and ocean-

induced wind systems (from the Atlantic and Indian Oceans), the seasonal precipitation at the fourteen grids exhibits high temporal and spatial variability. This can be illustrated by the monthly precipitation climatology presented in **Figure 5.2**. Notably, some of these monthly climatology series have dual peaks in the transition seasons (spring and fall), which are related to two major rainy seasons in East Africa: the "short-rains" and "long-rains" seasons.

The short rains, known as *Vuli* in Tanzania or *Deyr* in Somalia, are the boreal-autumn rains and prone to produce less rainfall but with larger interannual variability. In contrast, the long rains, known as *Masika* in Tanzania, *Belg* in Ethiopia, or *Gu* in Somalia, are the boreal-spring rains and generate more abundant rainfall with less interannual variability. In this study, the short rains period is defined as October–November (ON) for every grid examined [**Figure 5.3(a)**] except for Grid 6, in which the short rains period is November–December (ND) in accordance with the peak rainy period observed from the precipitation climatology. Likewise, the long rains period is defined as March–May (MAM) for every grid [**Figure 5.3(b)**] except for Grid 10, in which the long rains period is April–June (AMJ).

In addition to the short rains and long rains, the third rain type examined in this study is the "unimodal rains" in Tanzania (T-unimodal rains hereafter). The T-unimodal rains occur at three grids as shown in **Figure 5.3(c)**, and the rainy period is defined as November–April (NDJFMA) for Grids 1 and 2 and December–April (DJFMA) for Grid 5. The remaining rain types examined in this study include the "southern Sudan rains" and

the "*Kiremt* rains" in Ethiopia. The southern Sudan rains occur at Grid 4 with a rainy period spanning more than half a year (April–October, AMJJASO), while the Kiremt rains take place at Grids 9 and 12 with rainy periods in June–September (JJAS) and in July–September (JAS), respectively. The affected grids of these two rain types are shown in **Figure 5.3(d)**, and these two rain types may hereafter be referred to as the "other rains" in the following analysis.

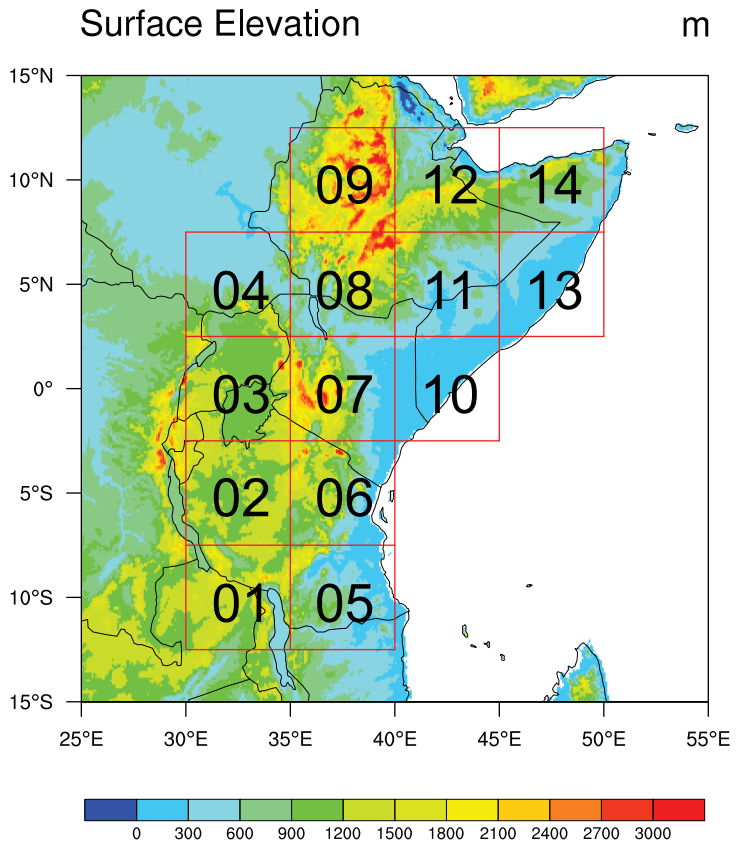



**Figure 5.1:** Fourteen precipitation grids based on aggregated CMAP data over East Africa.

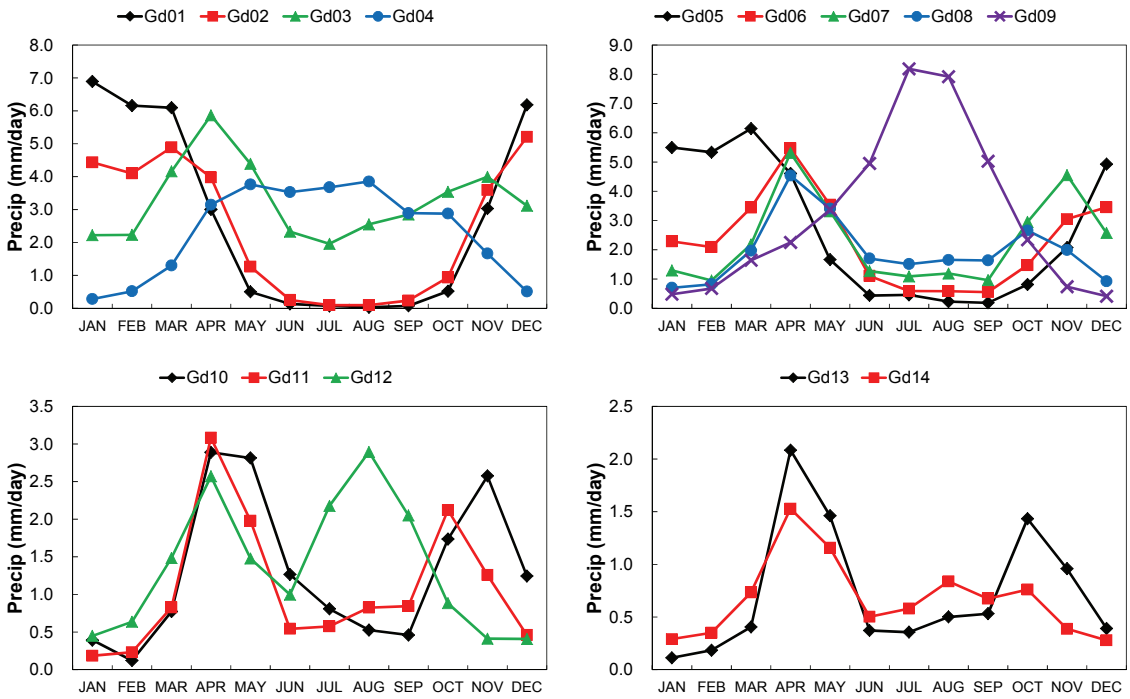



**Figure 5.2:** Monthly precipitation climatology corresponding to the fourteen grids delineated in **Figure 5.1**.

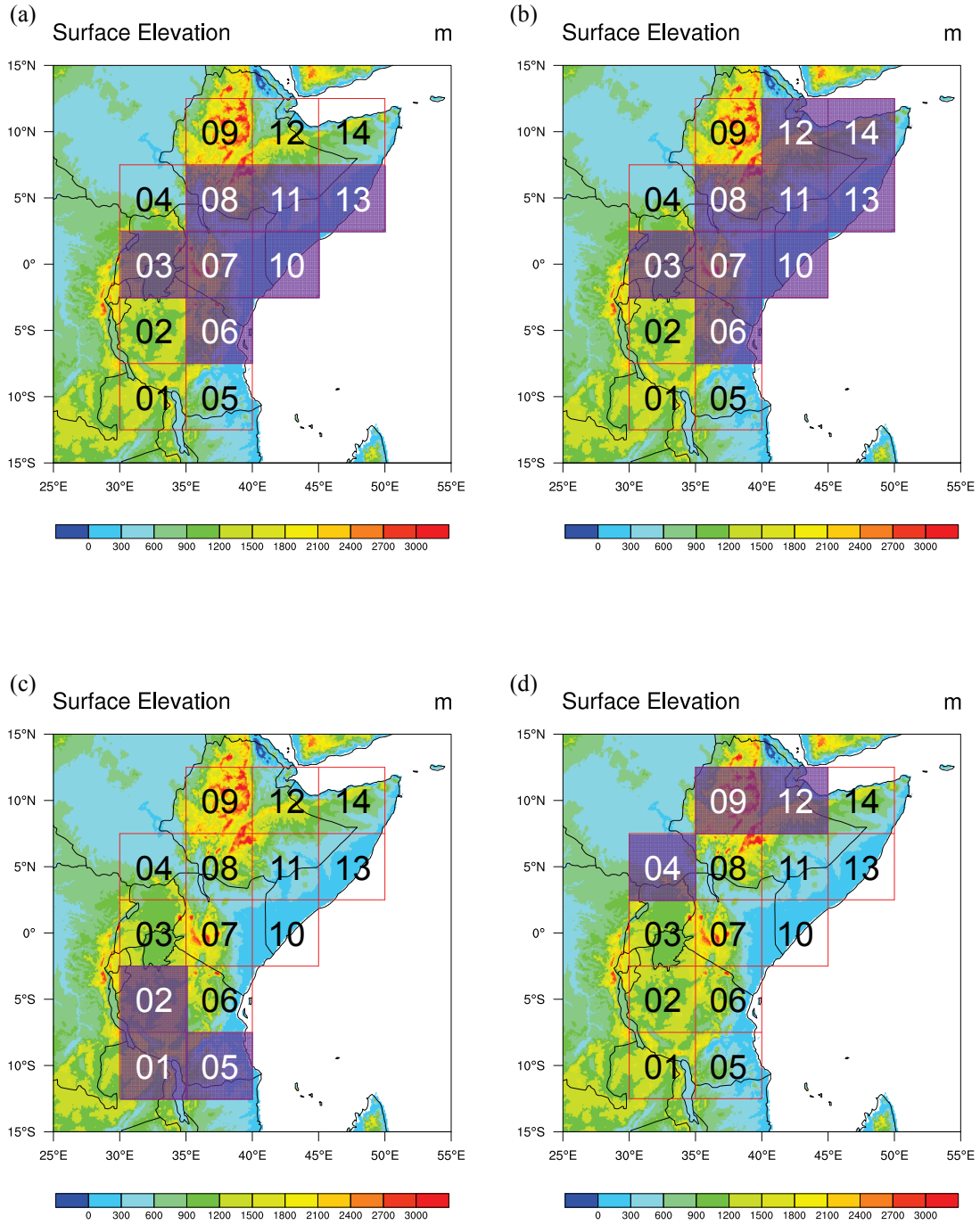



**Figure 5.3:** Grids affected by different rain types: (a) short rains, (b) long rains, (c) Tanzania unimodal rains, and (d) other rains (southern Sudan and Ethiopia Kiremt rains).

## 5.2 Seasonal Dipole Identification

In this section, the proposed dipole algorithm is applied to the hindcasting of four types of East African seasonal precipitation (i.e., the short rains, long rains, T-unimodal rains, and other rains) described in the last section. Precipitation series of these rain types are derived from the CMAP data from 1980 to 2009 as the target external series for forecasting.

To search SSTA dipole predictors for forecasting, the same Kaplan SST product is used. However, due to varied large-scale circulation mechanisms for the four rain types, different SSTA spatial domains are defined next. Because the variability of the East African short rains has been shown to strongly interact with the Indian Ocean (Hastenrath 2007; Ummenhofer et al., 2009), the Indian Ocean dipole (IOD) or zonal mode (Black, et al., 2003; Clark et al., 2003; Behera et al., 2005) in particular, an SSTA spatial domain (40°S–30°N, 40°E–130°E) that fully covers the equatorial Indian Ocean is defined for the dipole identification for the short rains [**Figure 5.4(a)**].

Regarding the long rains, although the understanding of mechanisms that dominate its variability has not well characterized yet, some studies suggested that Congo westerly (Nicholson, 1996; Pohl and Camberlin, 2006) and upper atmospheric anomalies over the Near East (Camberlin and Philippon, 2002) may be related to rainfall amounts in the long-rains season. Therefore, an SSTA spatial domain (40°S–60°N, 40°W–90°E) that covers not only the western Indian Ocean but also a large portion of the North and South Atlantic Oceans is defined for the dipole identification for the long rains [**Figure 5.4(b)**].

As for the T-unimodal and other rains, since limited studies have shown any plausible explanation related to SSTA or large-scale forcing, both SSTA spatial domains for the short and long rains are examined to explore higher predictability.

As the hindcasting experiment conducted in Chapter 4, the most skillful lead time for forecasting is found by executing the dipole algorithm repeatedly at various lead times. However, this procedure requires a few modifications because most of the rainy seasons in East Africa previously defined are different from the regular four seasons (i.e., DJF, MAM, JJA, and SON). For the sake of analytical consistency, each three-month average of SSTA data (i.e., DJF, JFM, …, NDJ) in the defined spatial domain is first evaluated so that the dipole algorithm can be executed repeatedly from 1- to 12-month lead times corresponding to any specific rain types. The monthly lead time still represents the difference between the first month of an SSTA season and that of a rainy season. For example, the SON SSTA has a 1-month lead time ahead of the ON short rains. The information of lead times, predictor domains, and other user-specified parameters for this study are summarized in **Table 5.1**.

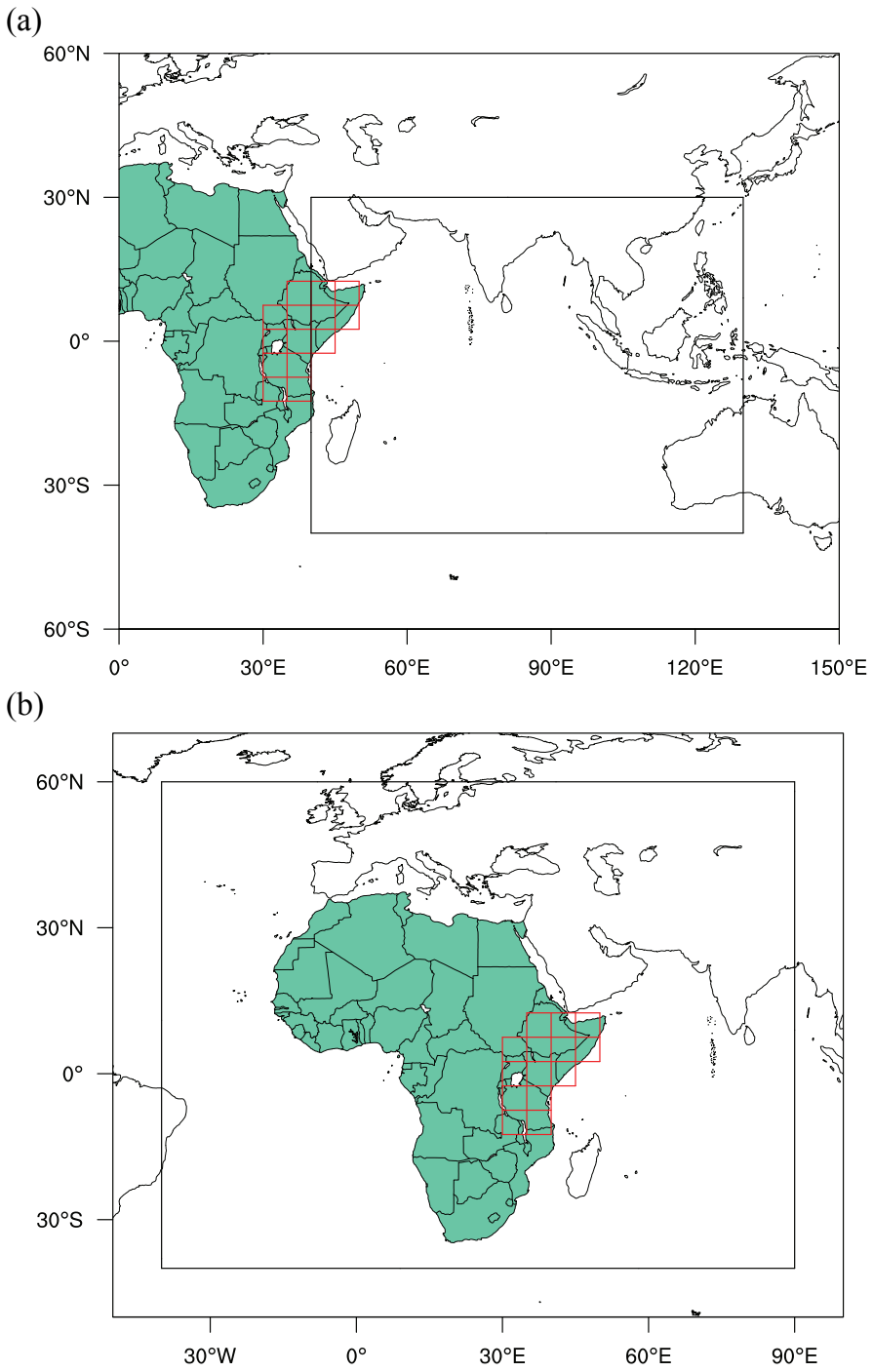



**Figure 5.4:** SSTA predictor domains used for the dipole identification for the (a) short rains and (b) long rains. Both domains are applied to the dipole identification for the T-unimodal and other rains.

**Table 5.1:** User-specified parameters for the case study in East Africa.

| **Data size** $(h)$ | **Predictor description** | **Predictand description** | **Upper threshold of clustering pixels** $[p \times q]$ | $GSS_{R,\alpha=0.05}$ | $\xi$ | **Lead times** |
|---|---|---|---|---|---|---|
| 30 [1980–2009, ON (ND for Grid 6)] | Kaplan SSTA [40°S–30°N, 40°E–130°E] | CMAP-aggregated East Africa short rains | $[p = q = 10]$ | 0.25 | 20 | 1, 2, …, 12 (months) |
| 30 [1980–2009, MAM (AMJ for Grid 10)] | Kaplan SSTA [40°S–60°N, 40°W–90°E] | CMAP-aggregated East Africa long rains | $[p = q = 10]$ | 0.25 | 20 | 1, 2, …, 12 (months) |
| 30 [1980–2009, NDJFMA (DJFMA for Grid 5)] | Kaplan SSTA [short and long rains domains] | CMAP-aggregated Tanzania unimodal rains | $[p = q = 10]$ | 0.25 | 20 | 1, 2, …, 12 (months) |
| 30 [1980–2009, AMJJASO, JJAS, and JAS] | Kaplan SSTA [short and long rains domains] | CMAP-aggregated southern Sudan and Kiremt rains | $[p = q = 10]$ | 0.25 | 20 | 1, 2, …, 12 (months) |

*5.2.1 Short Rains*

After the dipole algorithm is sequentially applied to the hindcasting of the short rains at the seven delineated grids [see **Figure 5.3(a)**], the subsequent forecasting skills in terms of two scalar accuracy measures (MAE and $\rho$) and four ensemble forecasting measures (Re, LME, UME, and $\gamma$) are recorded in **Table 5.2**. Although the dipole algorithm is sifted through various lead times, the forecasting skills for all grids agree that the best lead time for the East African short rains is only one month, implying strong concurrent interactions between short rains and SST forcing. The forecasting skill drops rapidly (especially for Grids 7, 8, and 10), leading to low long-lead skills for certain grids. Further, the algorithm sometimes cannot identify enough significant dipoles, indicating that an enlarged SSTA domain may be required to generate more dipole predictors.

Among the seven grids, the strongest predictability is exhibited at Grid 11 (over the border between Ethiopia, Somalia, and Kenya) with optimum MAE and Re values of 0.489 and 0.900, respectively. Strong predictability is also noted at Grid 3 (over Uganda and Lake Victoria) with an optimum Re value of 0.733 and best CV correlation coefficient of 0.714, explaining about 51% of the precipitation variance. In addition to these two grids, some weaker but still substantial skills can be found at other grids in which the Re values range from 0.600 to 0.733, and the best identified dipoles can account for 27–40% of the short rains variance at the 1-month lead time. However, if lead times sufficient for practical seasonal forecasting (i.e., at least three months or longer) are of interest, only dipoles identified for Grids 3, 10, and 11 are useful.

To express the correspondence between the ensemble forecasting and observation values at near-concurrent and longer lead times, hindcasting results at two different lead times are selected and plotted in **Figures 5.5** to **5.7** for all grids. The first selected lead time is one month for all grids due to consistent optimum skills as described, and the second selected lead time is at least three months or longer with relatively high skills for a certain grid. These figures not only reiterate the strong forecasting skills occurring at several grids (e.g., Grids 3 and 11) but also highlight an extreme flood event in 1997 that shows tremendous impacts on entire East Africa. Although nearly all forecasts are likely to respond to this event, the magnitude discrepancy affects the forecasting skills. Regardless of this probable outlier, the forecasting envelopes seem to be able to contain or respond to other events remarkably well, especially to droughts, explaining why LME values are generally better than UME values. This finding reaffirms the importance of incorporating the GSS in the dipole algorithm, which helps to establish a more reliable forecasting model less sensitive to any outliers. Nevertheless, if the rainfall is largely governed by chaotic atmosphere variations (e.g., rainfall at Grid 13), forecasting based on SSTA dipoles may have difficulty accounting for its variability.

**Figures 5.7** and **5.8** show the identified dipoles used to generate the hindcasts of the short rains at the seven grids. At the 1-month lead time, while active negative pole areas are mostly located at the eastern basin of the defined domain (from the vicinity of Indonesia and to as far north as the South China Sea), the active positive poles seem to disperse from the southern Indian Ocean to the western equatorial Indian Ocean. The zonal-oriented dipoles, distinguishable for Grids 10 and 11 (equatorial coastal regions),

illustrate their linkages to the IOD (e.g., Behera et al., 2005). The slight displacement of these dipoles compared to the ordinary IOD pattern represents a more enhanced predictor-predictand relationship introduced by the proposed algorithm. Thus, the algorithm is able to identify predictor regions resembling those with evidential physical explanation, but such regions are more conducive than those subjectively predefined regions to the short rains at the designated grids. In addition to the IOD-like dipoles, the contribution of other dipole components in the southern and western equatorial Indian Oceanic areas to the short rains are also supported by a recent study (Ummenhofer et al., 2009). However, their investigations, based on AGCM simulations with perturbed SST in prescribed areas, indicated that the anomalies in the western equatorial Indian Ocean can induce more significant changes in the short rains.

The identified dipoles at longer lead times present some migration of active oceanic areas. For example, the southwestern Indian Ocean near Madagascar seems to become crucial to the short rains at several grids (e.g., Grids 3, 6, 11, and 13) in one season ahead, possibly due to its relationship with the Indian Ocean trade winds and Southwest Monsoon. This also implies that the ordinary IOD pattern only carry little or no precursory information.

**Table 5.2:** Forecasting skill at different lead times for the short rains in East Africa; the best value of each measure is bold and underlined.

| Gd_03 | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | **0.523** | **0.714** | **0.733** | -0.938 | 1.695 | 0.371 |
| 2 | 0.554 | 0.678 | 0.633 | -0.888 | 1.712 | 0.345 |
| 3 | 0.573 | 0.658 | 0.633 | **-0.707** | 1.970 | 0.361 |
| 4 | 0.661 | 0.476 | 0.633 | -1.084 | 1.592 | 0.326 |
| 5 | 0.649 | 0.421 | 0.567 | -1.409 | 1.654 | **0.298** |
| 6 | 0.663 | 0.466 | 0.667 | -1.128 | 1.797 | 0.317 |
| 7 | 0.672 | 0.394 | 0.633 | -1.054 | 1.755 | 0.347 |
| 8 | 0.661 | 0.518 | 0.667 | -1.017 | **1.397** | 0.381 |
| 9 | 0.614 | 0.453 | 0.700 | -0.937 | 1.580 | 0.382 |
| 10 | 0.630 | 0.417 | 0.667 | -1.146 | 1.558 | 0.396 |
| 11 | 0.664 | 0.393 | 0.633 | -1.637 | 1.753 | 0.330 |
| 12 | 0.660 | 0.502 | 0.600 | -1.592 | 1.787 | 0.342 |
| **Gd_06** | | | | | | |
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | **0.827** | 0.560 | **0.633** | -1.263 | **1.515** | 0.367 |
| 2 | 0.830 | **0.603** | 0.500 | **-0.777** | 1.819 | 0.338 |
| 3 | 0.865 | 0.516 | 0.467 | -1.163 | 1.939 | 0.299 |
| 4 | 0.846 | 0.494 | 0.533 | -1.466 | 2.320 | 0.333 |
| 5 | 0.835 | 0.557 | 0.567 | -1.390 | 3.097 | 0.324 |
| 6 | 0.865 | 0.509 | 0.567 | -1.133 | 3.041 | 0.285 |
| 7 | 0.897 | 0.449 | 0.567 | -1.342 | 2.807 | **0.282** |
| 8 | 0.936 | 0.432 | 0.533 | -1.183 | 1.928 | 0.362 |
| 9 | 0.878 | 0.529 | 0.533 | -1.308 | 2.342 | 0.357 |
| 10 | 0.939 | 0.404 | 0.500 | -0.954 | 2.247 | 0.340 |
| 11 | 0.886 | 0.462 | 0.533 | -1.336 | 2.404 | 0.338 |
| 12 | 0.892 | 0.529 | 0.500 | -1.379 | 1.939 | 0.330 |

**Table 5.2:** *Continued.*

| Gd_07 | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | **0.965** | **0.607** | **0.733** | -0.719 | **4.352** | 0.303 |
| 2 | 1.061 | 0.237 | 0.533 | -1.371 | 5.113 | 0.251 |
| 3 | 1.111 | 0.098 | 0.500 | -1.208 | 5.345 | 0.203 |
| 4 | 1.121 | 0.076 | 0.500 | -1.201 | 5.393 | 0.204 |
| 5 | 1.156 | -0.090 | 0.233 | -1.283 | 5.750 | **0.126** |
| 6 | 1.116 | 0.111 | 0.400 | -1.356 | 5.722 | 0.146 |
| 7 | 1.087 | 0.328 | 0.467 | -0.806 | 5.707 | 0.196 |
| 8 | 1.068 | 0.417 | 0.667 | -0.860 | 5.466 | 0.226 |
| 9 | 1.013 | 0.472 | 0.567 | **-0.639** | 5.263 | 0.264 |
| 10 | 1.021 | 0.448 | 0.567 | -0.698 | 5.016 | 0.297 |
| 11 | 1.041 | 0.425 | 0.600 | -0.783 | 5.031 | 0.250 |
| 12 | 1.029 | 0.334 | 0.533 | -1.157 | 4.774 | 0.240 |
| **Gd_08** | | | | | | |
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ*** |
| 1 | **0.596** | **0.619** | **0.733** | **-0.195** | **2.928** | 0.313 |
| 2 | 0.713 | 0.323 | 0.400 | -0.646 | 3.670 | 0.229 |
| 3 | 0.710 | 0.036 | 0.400 | -0.743 | 4.007 | - |
| 4 | 0.753 | 0.095 | 0.167 | -1.024 | 4.309 | - |
| 5 | 0.744 | -0.374 | 0.167 | -0.961 | 4.126 | - |
| 6 | 0.686 | 0.006 | 0.400 | -0.543 | 4.174 | - |
| 7 | 0.652 | 0.302 | 0.500 | -0.425 | 3.569 | 0.239 |
| 8 | 0.651 | 0.321 | 0.600 | -0.487 | 3.249 | 0.249 |
| 9 | 0.626 | 0.500 | 0.633 | -0.347 | 3.456 | 0.247 |
| 10 | 0.616 | 0.333 | 0.633 | -0.329 | 3.904 | 0.243 |
| 11 | 0.676 | 0.275 | 0.433 | -0.596 | 3.975 | **0.188** |
| 12 | 0.681 | 0.295 | 0.533 | -0.656 | 3.977 | 0.216 |

*: "-" indicates the number of final dipoles is fewer than 20.

**Table 5.2:** *Continued*.

| Gd_10 | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ*** |
| 1 | 0.926 | 0.521 | **0.733** | **-0.147** | **3.489** | 0.333 |
| 2 | **0.897** | **0.584** | 0.700 | -0.530 | 4.400 | 0.313 |
| 3 | 0.951 | 0.527 | 0.700 | -1.227 | 4.498 | 0.299 |
| 4 | 1.025 | 0.312 | 0.667 | -1.211 | 4.949 | 0.270 |
| 5 | 1.031 | 0.174 | 0.500 | -1.224 | 5.529 | 0.215 |
| 6 | 1.058 | 0.174 | 0.400 | -0.831 | 5.585 | **0.212** |
| 7 | 1.022 | 0.171 | 0.667 | -0.294 | 4.759 | 0.285 |
| 8 | 1.047 | 0.324 | 0.633 | -0.761 | 5.296 | 0.251 |
| 9 | 1.111 | 0.108 | 0.367 | -0.905 | 5.742 | - |
| 10 | 1.065 | 0.158 | 0.433 | -1.045 | 5.522 | - |
| 11 | 1.078 | 0.140 | 0.567 | -0.740 | 4.957 | 0.246 |
| 12 | 1.015 | 0.246 | 0.533 | -0.913 | 4.917 | 0.276 |
| **Gd_11** | | | | | | |
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | **0.489** | 0.674 | **0.900** | **-0.322** | **2.030** | 0.320 |
| 2 | 0.536 | **0.707** | 0.767 | -0.660 | 2.496 | 0.293 |
| 3 | 0.529 | 0.580 | 0.733 | -0.758 | 2.795 | 0.272 |
| 4 | 0.651 | 0.251 | 0.533 | -0.743 | 3.729 | 0.248 |
| 5 | 0.696 | -0.050 | 0.333 | -0.963 | 4.008 | **0.155** |
| 6 | 0.674 | 0.299 | 0.400 | -0.812 | 3.994 | 0.185 |
| 7 | 0.649 | 0.203 | 0.567 | -0.587 | 3.434 | 0.242 |
| 8 | 0.662 | 0.461 | 0.500 | -0.395 | 3.473 | 0.228 |
| 9 | 0.668 | 0.301 | 0.667 | -0.480 | 3.106 | 0.260 |
| 10 | 0.648 | 0.307 | 0.633 | -0.384 | 3.416 | 0.258 |
| 11 | 0.655 | 0.187 | 0.500 | -0.900 | 3.677 | 0.218 |
| 12 | 0.692 | 0.093 | 0.533 | -0.828 | 3.297 | 0.219 |

*: "-" indicates the number of final dipoles is fewer than 20.

**Table 5.2:** *Continued*.

| Gd_13 | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.573 | **0.627** | **0.600** | -0.455 | **1.188** | 0.409 |
| 2 | 0.558 | 0.585 | 0.567 | -0.488 | 1.514 | 0.380 |
| 3 | 0.571 | 0.563 | 0.533 | **-0.408** | 1.538 | 0.373 |
| 4 | 0.606 | 0.361 | 0.400 | -0.646 | 1.866 | 0.320 |
| 5 | 0.643 | 0.331 | 0.300 | -0.666 | 1.796 | **0.274** |
| 6 | 0.573 | 0.445 | 0.433 | -0.568 | 1.968 | 0.311 |
| 7 | **0.554** | 0.590 | 0.433 | -0.647 | 1.843 | 0.330 |
| 8 | 0.606 | 0.306 | 0.433 | -0.551 | 1.814 | 0.362 |
| 9 | 0.603 | 0.367 | 0.467 | -0.477 | 1.729 | 0.339 |
| 10 | 0.571 | 0.535 | 0.500 | -0.460 | 1.426 | 0.383 |
| 11 | 0.637 | 0.394 | 0.433 | -0.824 | 1.570 | 0.355 |
| 12 | 0.619 | 0.461 | 0.467 | -0.622 | 1.513 | 0.330 |

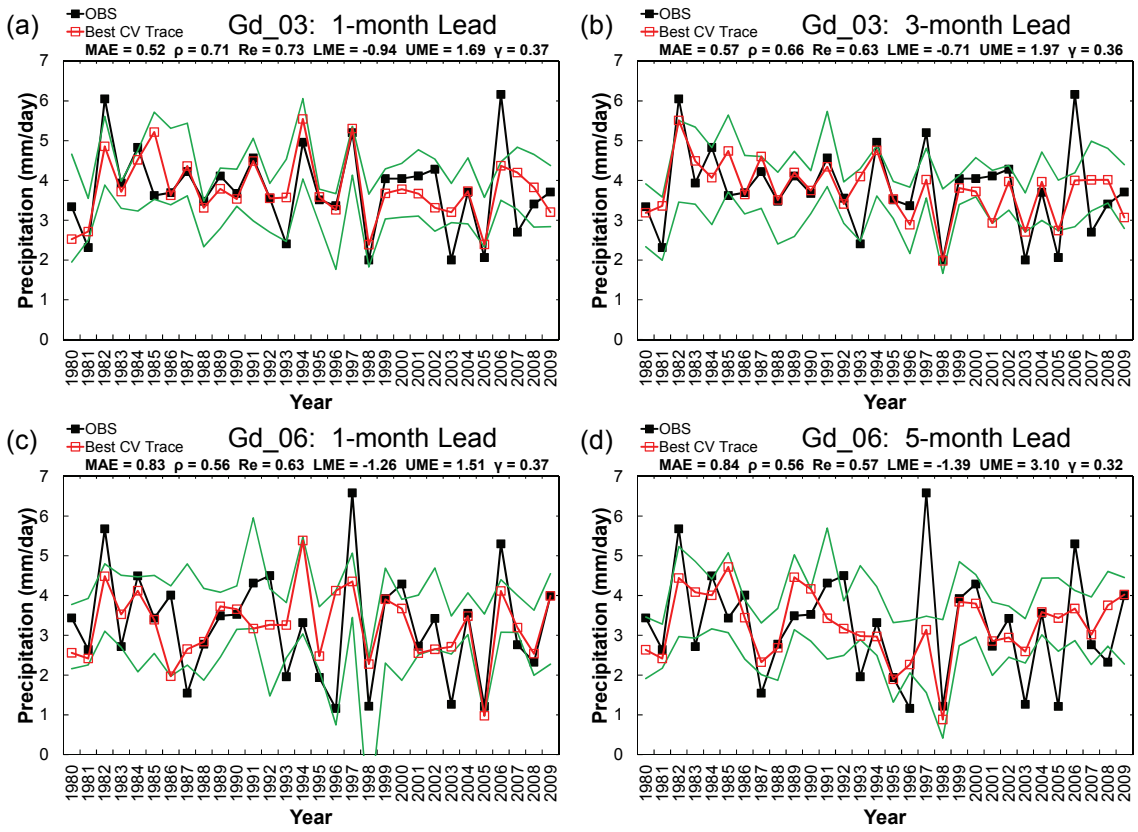



**Figure 5.5:** Optimal hindcasting results at two selected lead times for the short rains at Grids 3 and 6. The best CV trace is shown in red, and the green envelope is constructed using the maximum and minimum hindcasting values.

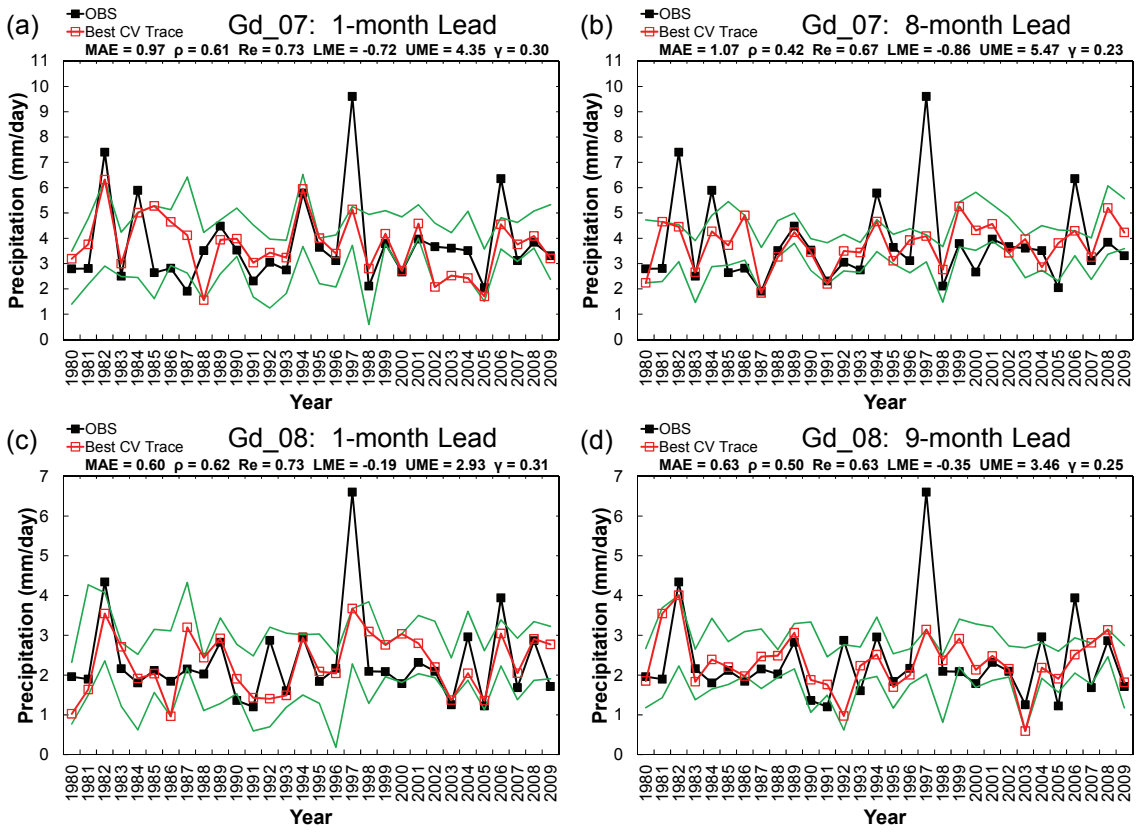



**Figure 5.6:** As in **Figure 5.5**, but for Grids 7 and 8.

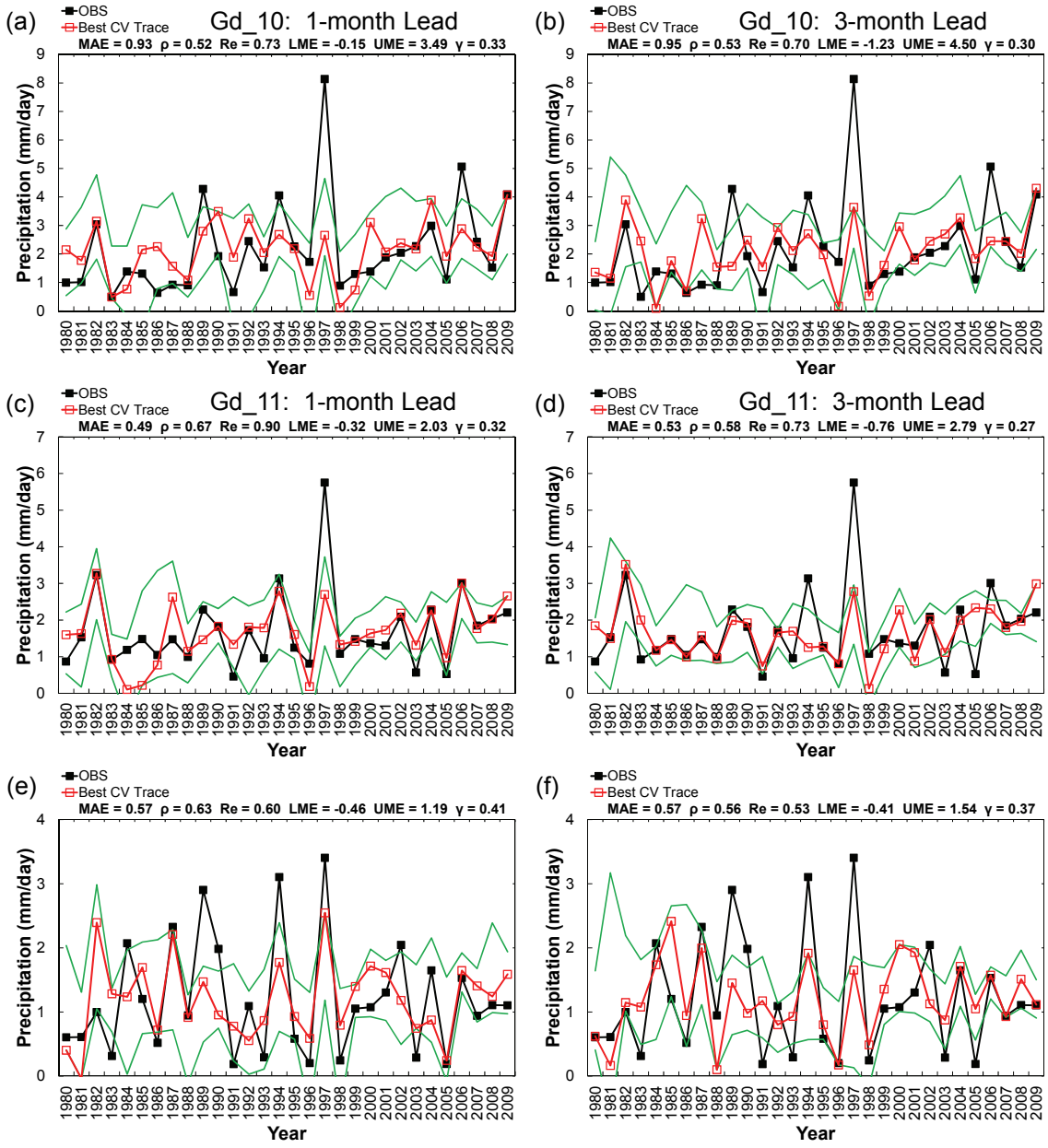



**Figure 5.7:** As in **Figure 5.5**, but for Grids 10, 11, and 13.

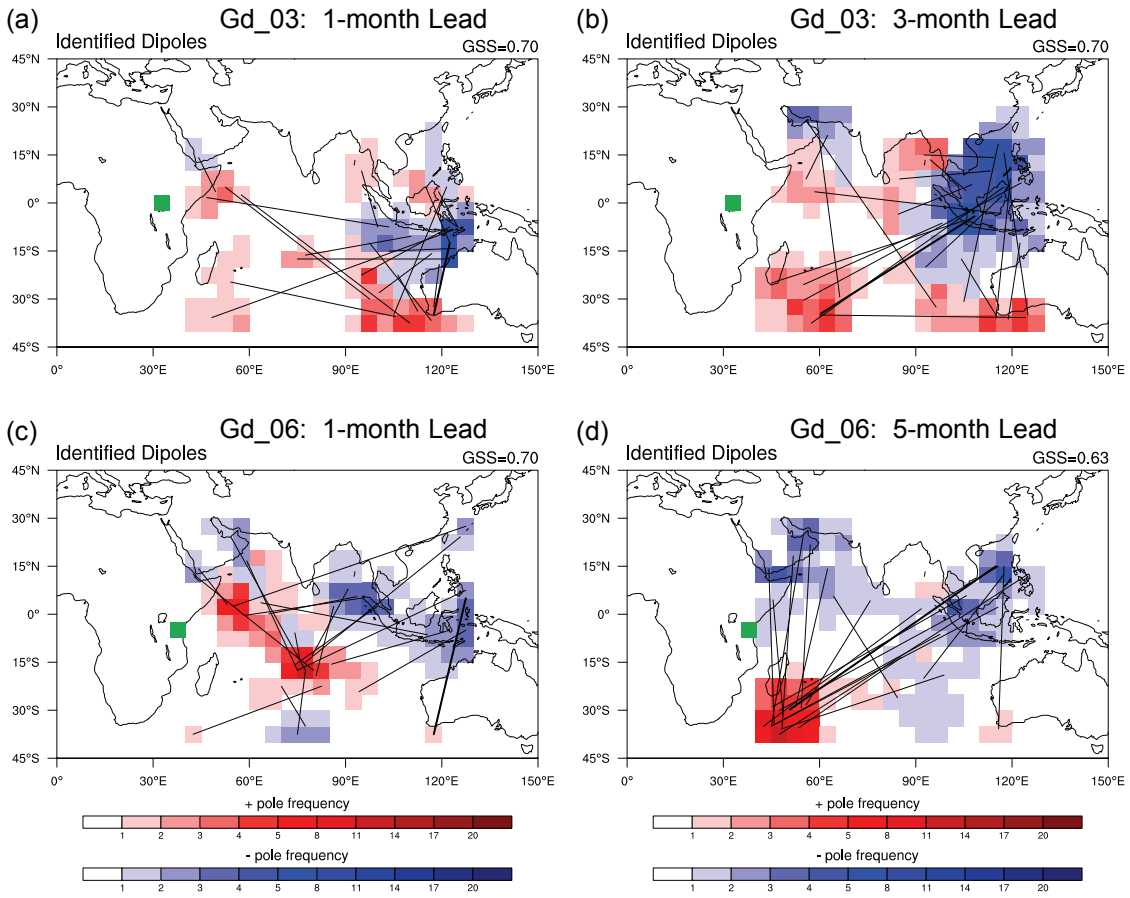



**Figure 5.8:** Corresponding dipoles used to generate the hindcasting results shown in **Figure 5.5**. Grid locations are denoted in green cells. Thicker and thinner connective lines indicate the orientation of the best and remaining dipoles, respectively. The best GSS among all dipoles obtained during the predictor screening process is denoted at the upper-right corner of each figure.

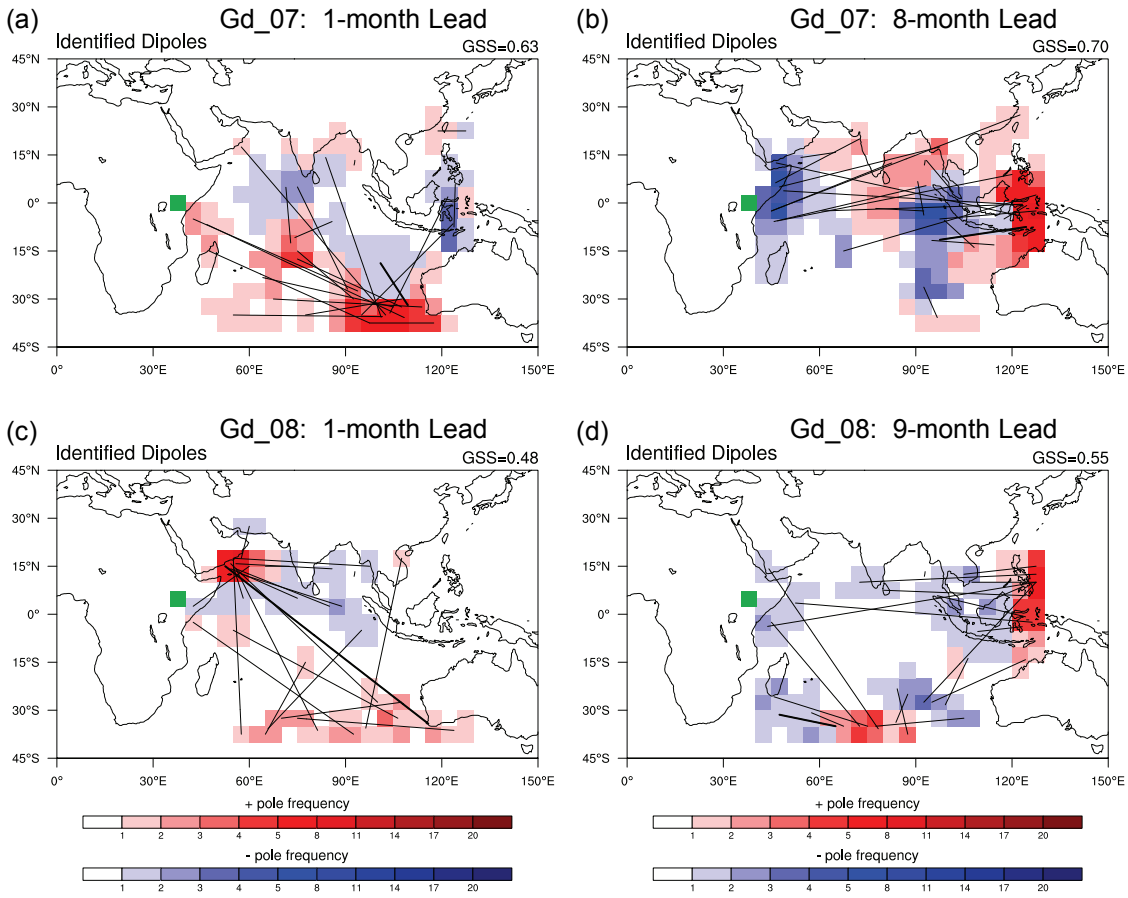



**Figure 5.9:** As in **Figure 5.8**, but for the hindcasting results shown in **Figure 5.6**.

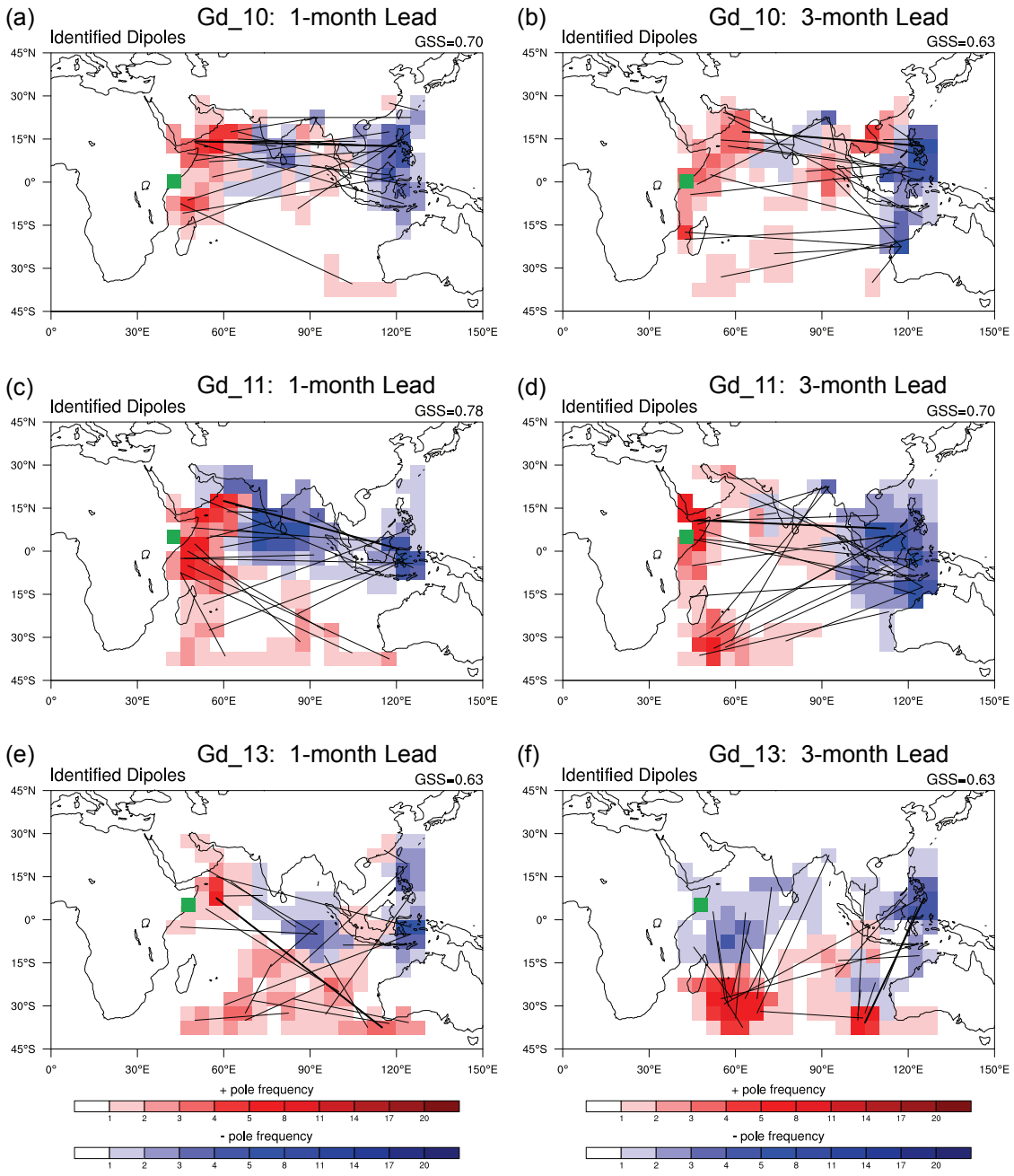



**Figure 5.10:** As in **Figure 5.8**, but for the hindcasting results shown in **Figure 5.7**.

*5.2.2 Long Rains*

The second rain type to be analyzed is the long rains at the nine delineated grids [see **Figure 5.3(b)**]. The associated forecasting skill at different lead times is recorded in **Table 5.3**. The strongest predictability appears at Grid 3 with the highest Re value of 0.800 (at the 11-month lead time), and the best CV correlation coefficient is 0.758 (at the 9-month lead time), explaining about 57% of precipitation variance. Because both the short- and long-rains seasons present satisfactory hindcasting results, the potential utility of the forecasts is expected to be more significant for Grid 3. Strong predictability is also noted at Grid 7 with an Re value of 0.767 and best CV correlation coefficient of 0.709 at the 11-month lead time. Other significant forecasting skill (both Re and $\rho$ values are greater than 0.6) is displayed at Grids 8, 10, 11, and 13. The weakest predictability is shown at Grid 12 (northeastern Ethiopia) with low Re and $\rho$ values of 0.533 and 0.494. Unlike the results for the short rains, the optimal lead times for forecasting the long rains are quite diverse and always beyond the near-concurrent period. For instance, while the long lead times have already been shown for Grids 3 and 10 (also at other grids), one to two seasons of lead times (i.e., three to six months) can be observed at some other grids, such as Grids 11, 13, and 14.

Accordingly, hindcasting results at two different lead times that exhibit the optimum and secondary forecasting skills (mainly judged by Re values) are selected and plotted in **Figures 5.11** to **5.13** for all grids. In addition to the lower interannual variability mentioned in Section 5.1, the observations of the long rains at some grids (e.g., Grids 3, 6, 7, and 8) display a prominent downward trend in the latest 30 years, which may be

related to global warming or the abrupt change in a large-scale precipitation pattern found in recent studies (e.g., Williams and Funk, 2010; Lyon and DeWitt, 2012). The interior grids generally receive higher rainfall amounts than the coastal grids because of a local topographic effect and moisture bearing westerly advection. In certain years, the average rain rates during the long-rains season over Somalia (*Gu*, at Grids 13 and 14) were near zero, which threatened local agriculture and famine. Regarding the forecasting results, in general, better correspondence between the forecasting envelopes and observations can be found at grids presenting a downward trend. Further, the identified dipoles seem to have minor problems in predicting dry than wet events, implying that other atmospheric mechanisms, such as the Madden-Julian Oscillation (Pohl and Camberlin, 2006), are responsible for the residual variability of the long rains (e.g., onset, cessation, and intensity). Thus, these mixed results (i.e., the spatial-varied observations, the diverse optimal lead times, and the partially explained variability) all signify the potential randomness of the long rains and its uncertain relationship with oceanic forcing.

One way to examine if the hindcasting skill is generated by statistical randomness is through the observation of the corresponding dipoles, as plotted in **Figures 5.14** to **5.16**. The identified dipoles for Grid 3 display organized, meridional dipole and even tripole patterns located at concentrated oceanic areas, such as the South Atlantic Ocean, the Mediterranean Sea, and the southwestern Indian Ocean. From the 11- to 5-month lead times, a smooth transition of dipole patterns (e.g., both the positive and negative poles migrate slightly eastward) can be observed, which mimics the real evolution of the SST field thereby supporting the physical significance of the identified predictor structures.

For Grid 6, the aforementioned oceanic areas remain active, yet the scattered dipole patterns indicate that the hindcasting skill may deteriorate. For Grid 7, even though it is adjacent to Grid 3, the identified dipoles at the same 11-month lead time reveal very distinct oceanic areas: The Arabian Sea and the East Coast of Africa become the main components of the positive and negative poles, respectively. Nevertheless, if the dipoles are as scattered as the patterns shown for Grid 7 at the 4-month lead time, the physical significance of the associated skill may be questionable; thus, these types of dipoles could be eliminated in operational forecasting practice.

The identified dipoles for Grids 8 and 10 at the 12-month lead time appear to inherit some of the main components seen for Grid 6 with additional active oceanic areas, including the Northeast Coast of Africa for Grid 8 and the South Atlantic Ocean near Brazil and the South Coast of Africa for Grid 10. The latter oceanic areas also constitute the main components of dipoles for Grid 11 at the 9-month lead time. However, at short to near-concurrent lead times, while the dipole patterns identified for Grid 10 are still prominent and indicate a concentrated negative pole at the North Atlantic, those for Grids 8 and 11 are quite scattered. As approaching to grids near the Horn of Africa regions (Grids 12, 13, and 14), the dipoles seem to be more and more scattered again at the short lead times. Among these scattered patterns, the more concentrated poles can only be seen in certain oceanic areas, such as the South Atlantic and the Central to South Indian Oceans. Overall, the dipole patterns present large variations from grid to grid, but some connections between patterns in adjacent grids can still be observed. Nonetheless, most of the dipole patterns presented here are unprecedented (due to lack of related literature

on long rains), so further investigations are necessary to provide more physical justifications for these patterns.

**Table 5.3:** As in **Table 5.2**, but for the long rains in East Africa.

| Gd_03 | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.521 | 0.605 | 0.700 | -0.866 | 1.136 | 0.362 |
| 2 | 0.531 | 0.550 | 0.700 | -1.187 | 1.080 | 0.349 |
| 3 | 0.561 | 0.411 | 0.600 | -1.085 | 1.334 | **0.328** |
| 4 | 0.531 | 0.592 | 0.667 | -1.016 | 1.538 | 0.335 |
| 5 | 0.515 | 0.684 | 0.733 | -0.597 | 0.699 | 0.366 |
| 6 | 0.519 | 0.491 | 0.667 | -0.877 | 0.827 | 0.347 |
| 7 | 0.530 | 0.636 | 0.700 | -0.862 | 0.966 | 0.356 |
| 8 | 0.486 | 0.632 | 0.667 | -0.815 | **0.422** | 0.358 |
| 9 | **0.443** | **0.758** | 0.767 | **-0.138** | 0.657 | 0.386 |
| 10 | 0.451 | 0.708 | 0.767 | -0.823 | 0.776 | 0.369 |
| 11 | 0.476 | 0.686 | **0.800** | -0.822 | 0.864 | 0.366 |
| 12 | 0.506 | 0.623 | 0.667 | -1.171 | 0.909 | 0.349 |
| **Gd_06** | | | | | | |
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.717 | 0.423 | 0.433 | -1.493 | 0.794 | **0.331** |
| 2 | 0.708 | 0.415 | 0.467 | -1.563 | 0.901 | 0.340 |
| 3 | 0.746 | 0.436 | 0.433 | -1.543 | 0.791 | 0.362 |
| 4 | 0.638 | **0.611** | 0.533 | -1.241 | 0.917 | 0.411 |
| 5 | 0.626 | 0.598 | 0.567 | -1.257 | 0.609 | 0.390 |
| 6 | 0.651 | 0.486 | **0.600** | -1.342 | 0.730 | 0.419 |
| 7 | 0.627 | 0.564 | **0.600** | -1.613 | 0.615 | 0.426 |
| 8 | 0.640 | 0.582 | 0.567 | -1.136 | 0.389 | 0.448 |
| 9 | **0.617** | 0.578 | 0.533 | -1.039 | 0.432 | 0.435 |
| 10 | 0.685 | 0.513 | 0.467 | **-0.889** | 0.712 | 0.390 |
| 11 | 0.692 | 0.506 | 0.467 | -1.402 | 0.907 | 0.402 |
| 12 | 0.644 | 0.546 | 0.567 | -1.388 | **0.234** | 0.414 |

**Table 5.3:** *Continued*.

| Gd_07 | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.649 | 0.573 | 0.600 | -1.333 | 1.588 | 0.298 |
| 2 | 0.691 | 0.545 | 0.567 | -1.436 | 2.262 | 0.269 |
| 3 | 0.641 | 0.563 | 0.633 | -1.460 | 2.053 | 0.286 |
| 4 | 0.652 | 0.400 | 0.667 | **-0.810** | 2.237 | 0.259 |
| 5 | 0.760 | 0.324 | 0.333 | -1.406 | 2.335 | 0.214 |
| 6 | 0.815 | 0.085 | 0.133 | -2.142 | 2.801 | **0.173** |
| 7 | 0.703 | 0.422 | 0.400 | -1.581 | 2.576 | 0.243 |
| 8 | 0.630 | 0.534 | 0.600 | -1.562 | 2.165 | 0.287 |
| 9 | 0.677 | 0.461 | 0.567 | -1.389 | 1.531 | 0.301 |
| 10 | 0.671 | 0.413 | 0.600 | -1.402 | 1.246 | 0.321 |
| 11 | **0.566** | **0.709** | **0.767** | -1.160 | **0.577** | 0.322 |
| 12 | 0.583 | 0.681 | 0.700 | -1.639 | 0.812 | 0.298 |
| **Gd_08** | | | | | | |
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | **0.470** | **0.673** | 0.700 | **-0.641** | 0.846 | 0.317 |
| 2 | 0.480 | 0.670 | 0.567 | -0.769 | 0.735 | 0.317 |
| 3 | 0.513 | 0.603 | 0.567 | -0.987 | 0.915 | 0.321 |
| 4 | 0.559 | 0.502 | 0.600 | -1.194 | 1.393 | 0.303 |
| 5 | 0.562 | 0.356 | 0.567 | -1.146 | 1.383 | 0.298 |
| 6 | 0.563 | 0.359 | 0.500 | -1.130 | 1.509 | **0.278** |
| 7 | 0.580 | 0.571 | 0.467 | -0.748 | 1.342 | 0.305 |
| 8 | 0.599 | 0.394 | 0.567 | -1.141 | 1.343 | 0.322 |
| 9 | 0.578 | 0.524 | 0.600 | -1.135 | 1.114 | 0.344 |
| 10 | 0.539 | 0.463 | 0.667 | -1.169 | 0.802 | 0.356 |
| 11 | 0.508 | 0.616 | **0.733** | -1.190 | 0.752 | 0.349 |
| 12 | 0.493 | 0.619 | **0.733** | -0.965 | **0.623** | 0.368 |

**Table 5.3:** *Continued.*

| Gd_10 | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.499 | **0.605** | 0.667 | -0.756 | 0.860 | 0.370 |
| 2 | 0.460 | 0.574 | 0.567 | -0.527 | 0.891 | 0.373 |
| 3 | 0.488 | 0.544 | 0.600 | -0.630 | **0.709** | 0.366 |
| 4 | 0.534 | 0.457 | 0.667 | -0.674 | 1.152 | 0.391 |
| 5 | 0.544 | 0.393 | 0.533 | -1.055 | 1.126 | 0.327 |
| 6 | 0.576 | 0.146 | 0.367 | -0.778 | 1.275 | 0.322 |
| 7 | 0.575 | 0.238 | 0.433 | -0.955 | 1.259 | 0.312 |
| 8 | 0.593 | 0.197 | 0.367 | -1.108 | 1.131 | **0.292** |
| 9 | 0.583 | 0.290 | 0.400 | -1.064 | 1.173 | 0.316 |
| 10 | 0.561 | 0.400 | 0.467 | -1.085 | 1.145 | 0.359 |
| 11 | 0.539 | 0.351 | 0.433 | -0.738 | 1.278 | 0.336 |
| 12 | **0.445** | 0.602 | **0.700** | **-0.334** | 1.177 | 0.420 |
| **Gd_11** | | | | | | |
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.431 | 0.426 | 0.533 | -0.679 | 1.067 | 0.321 |
| 2 | **0.363** | 0.510 | 0.600 | -0.601 | 0.925 | 0.329 |
| 3 | 0.392 | **0.614** | **0.667** | -0.551 | **0.653** | 0.376 |
| 4 | 0.419 | 0.417 | 0.567 | -0.446 | 0.699 | 0.365 |
| 5 | 0.429 | 0.475 | 0.567 | -0.370 | 1.065 | 0.346 |
| 6 | 0.428 | 0.379 | 0.533 | -0.484 | 1.137 | 0.315 |
| 7 | 0.434 | 0.345 | 0.433 | -0.533 | 1.449 | 0.282 |
| 8 | 0.439 | 0.363 | 0.367 | -0.616 | 1.472 | **0.265** |
| 9 | 0.419 | 0.410 | 0.600 | -0.390 | 1.096 | 0.343 |
| 10 | 0.412 | 0.464 | 0.567 | **-0.351** | 0.864 | 0.364 |
| 11 | 0.396 | 0.465 | 0.533 | -0.529 | 0.779 | 0.336 |
| 12 | 0.390 | 0.501 | 0.567 | -0.541 | 1.202 | 0.295 |

**Table 5.3:** *Continued.*

| Gd_12 | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.511 | 0.167 | 0.300 | -0.742 | 1.595 | 0.239 |
| 2 | 0.521 | 0.241 | 0.333 | -0.831 | 1.554 | **0.224** |
| 3 | 0.486 | 0.219 | 0.500 | -0.551 | 1.389 | 0.269 |
| 4 | 0.499 | 0.355 | 0.433 | -0.573 | 1.348 | 0.308 |
| 5 | **0.472** | **0.494** | **0.533** | **-0.516** | 1.349 | 0.335 |
| 6 | 0.479 | 0.355 | 0.433 | -0.605 | 1.477 | 0.315 |
| 7 | 0.532 | 0.121 | 0.367 | -0.650 | 1.489 | 0.268 |
| 8 | 0.517 | 0.277 | 0.333 | -1.004 | 1.441 | 0.251 |
| 9 | 0.513 | 0.140 | 0.333 | -0.886 | 1.380 | 0.244 |
| 10 | 0.550 | 0.089 | 0.300 | -0.833 | 1.502 | 0.243 |
| 11 | 0.505 | 0.253 | 0.400 | -0.711 | 1.366 | 0.283 |
| 12 | 0.551 | 0.138 | 0.300 | -0.694 | **1.178** | 0.286 |
| **Gd_13** | | | | | | |
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.481 | 0.346 | 0.667 | -0.859 | 1.950 | 0.269 |
| 2 | 0.458 | 0.361 | 0.600 | -0.893 | 2.019 | 0.249 |
| 3 | **0.419** | 0.553 | **0.733** | -0.697 | 1.640 | 0.270 |
| 4 | 0.432 | 0.505 | 0.700 | -0.810 | 1.762 | 0.243 |
| 5 | 0.459 | 0.432 | 0.700 | -0.643 | 1.623 | 0.273 |
| 6 | 0.435 | **0.601** | 0.700 | -0.943 | **1.563** | 0.271 |
| 7 | 0.465 | 0.450 | 0.533 | -1.041 | 1.945 | 0.224 |
| 8 | 0.456 | 0.348 | 0.600 | -0.777 | 1.961 | 0.225 |
| 9 | 0.473 | 0.343 | 0.600 | -0.744 | 2.025 | **0.198** |
| 10 | 0.433 | 0.536 | 0.600 | **-0.359** | 1.812 | 0.286 |
| 11 | 0.475 | 0.463 | 0.633 | -0.649 | 1.701 | 0.293 |
| 12 | 0.464 | 0.455 | 0.633 | -0.733 | 1.925 | 0.240 |

**Table 5.3:** *Continued*.

| Gd_14 | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.481 | 0.267 | 0.333 | -0.810 | 1.529 | **0.251** |
| 2 | 0.467 | 0.420 | 0.400 | -0.905 | 1.575 | 0.255 |
| 3 | 0.437 | 0.482 | 0.567 | -0.549 | 1.219 | 0.313 |
| 4 | **0.436** | 0.512 | 0.500 | -0.492 | 1.421 | 0.319 |
| 5 | 0.447 | **0.539** | **0.633** | **-0.286** | 1.473 | 0.338 |
| 6 | 0.465 | 0.423 | **0.633** | -0.532 | 1.224 | 0.379 |
| 7 | 0.483 | 0.408 | 0.600 | -0.596 | **1.115** | 0.337 |
| 8 | 0.473 | 0.383 | 0.533 | -0.737 | 1.120 | 0.309 |
| 9 | 0.500 | 0.358 | 0.400 | -0.851 | 1.503 | 0.315 |
| 10 | 0.459 | 0.380 | 0.433 | -0.677 | 1.358 | 0.335 |
| 11 | 0.505 | 0.256 | 0.500 | -0.529 | 1.189 | 0.351 |
| 12 | 0.484 | 0.477 | 0.500 | -0.706 | 1.458 | 0.332 |

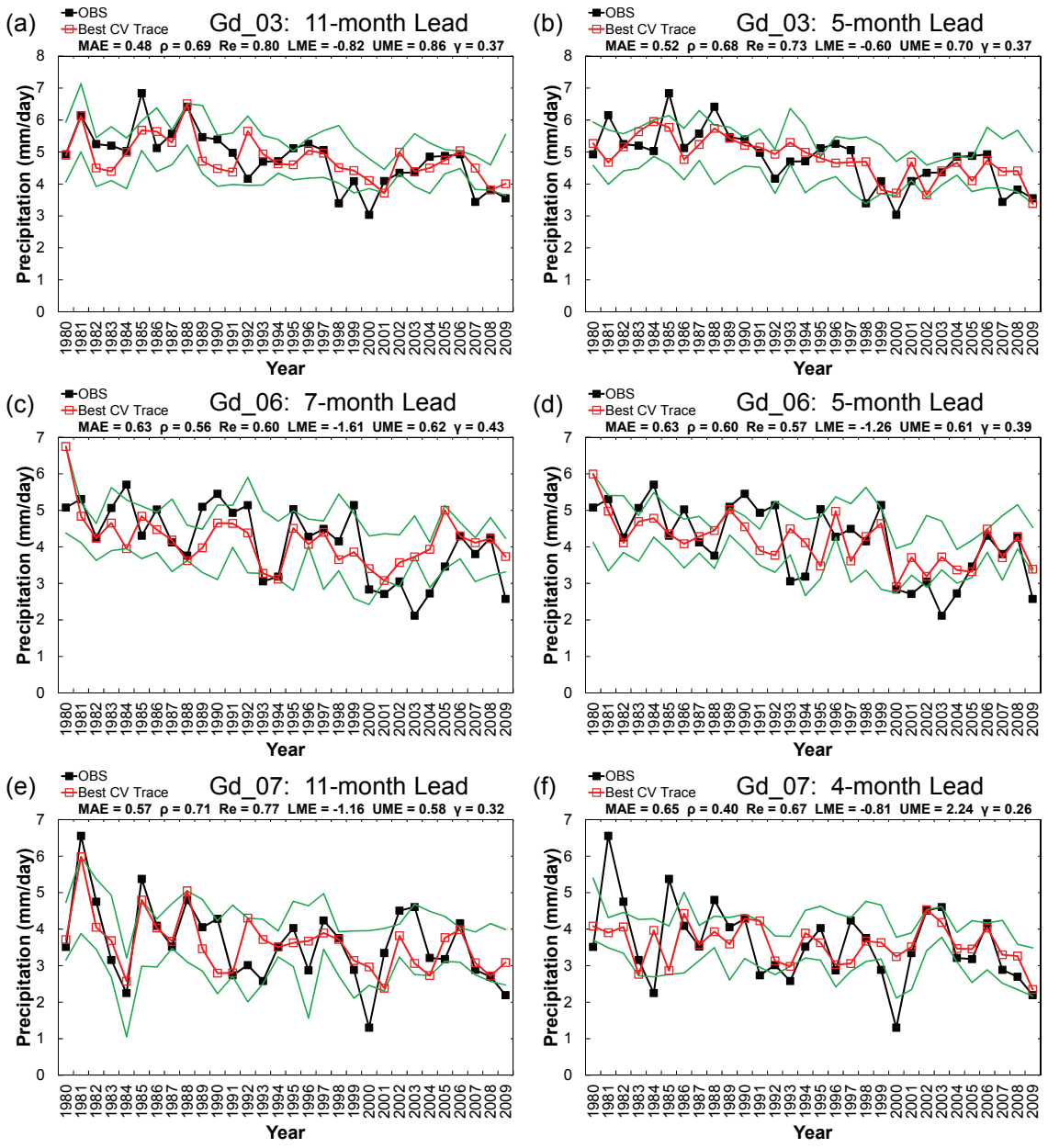



**Figure 5.11:** Optimal hindcasting results at two selected lead times for the long rains at Grids 3, 6, and 7. The best CV trace is shown in red, and the green envelope is constructed using the maximum and minimum hindcasting values.

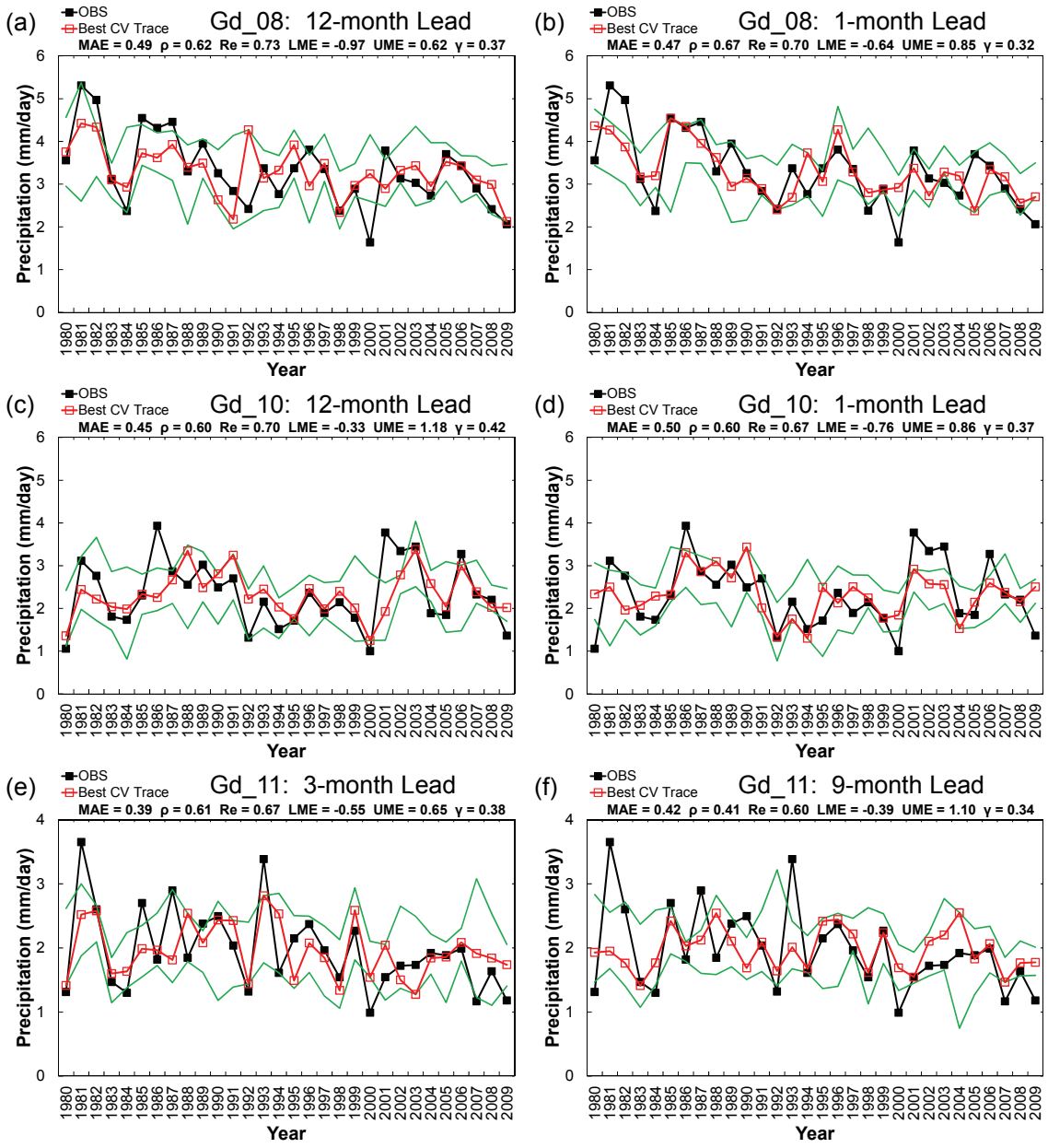



**Figure 5.12:** As in **Figure 5.11**, but for Grids 8, 10, and 11.

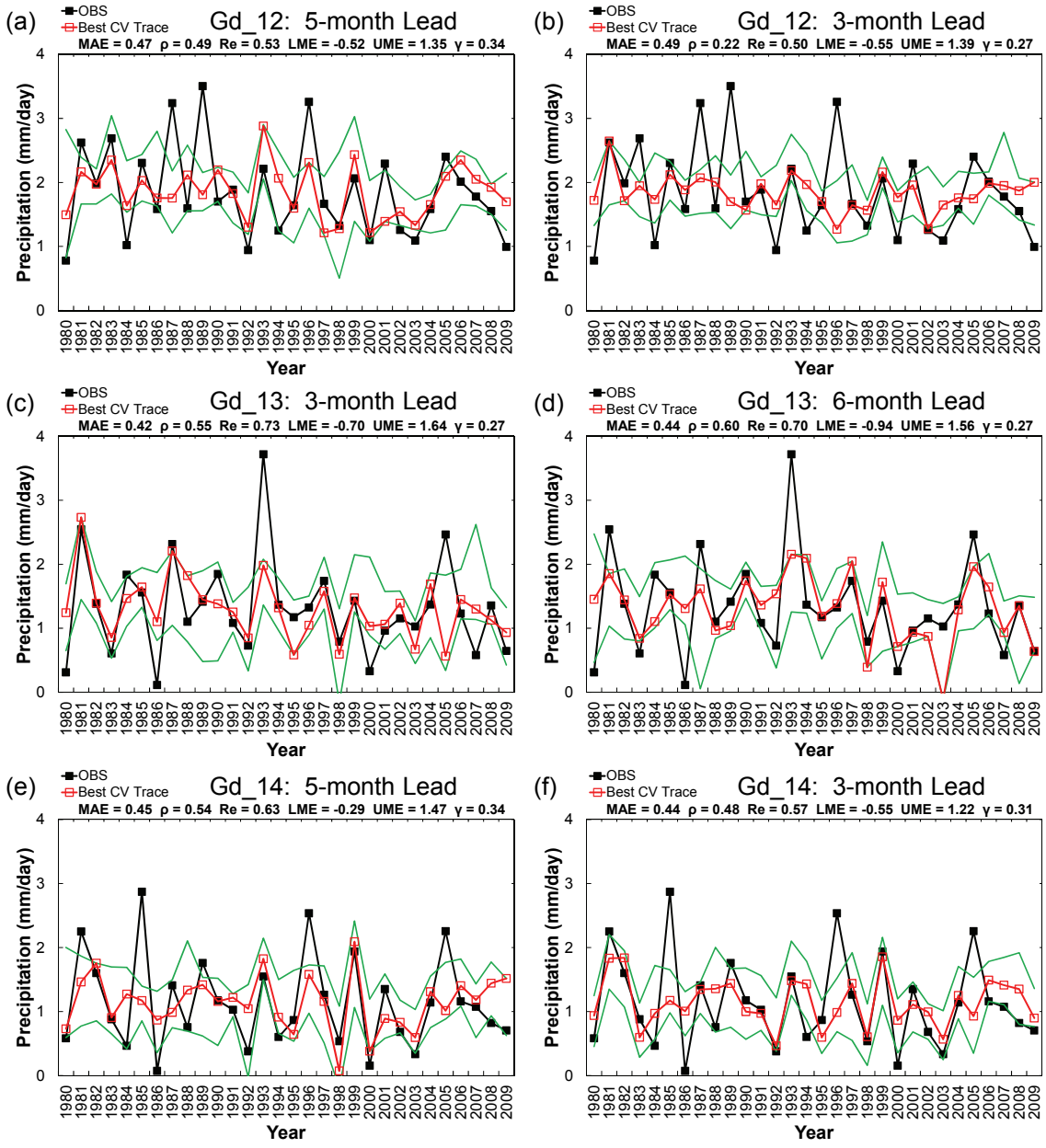



**Figure 5.13:** As in **Figure 5.11**, but for Grids 12, 13, and 14.

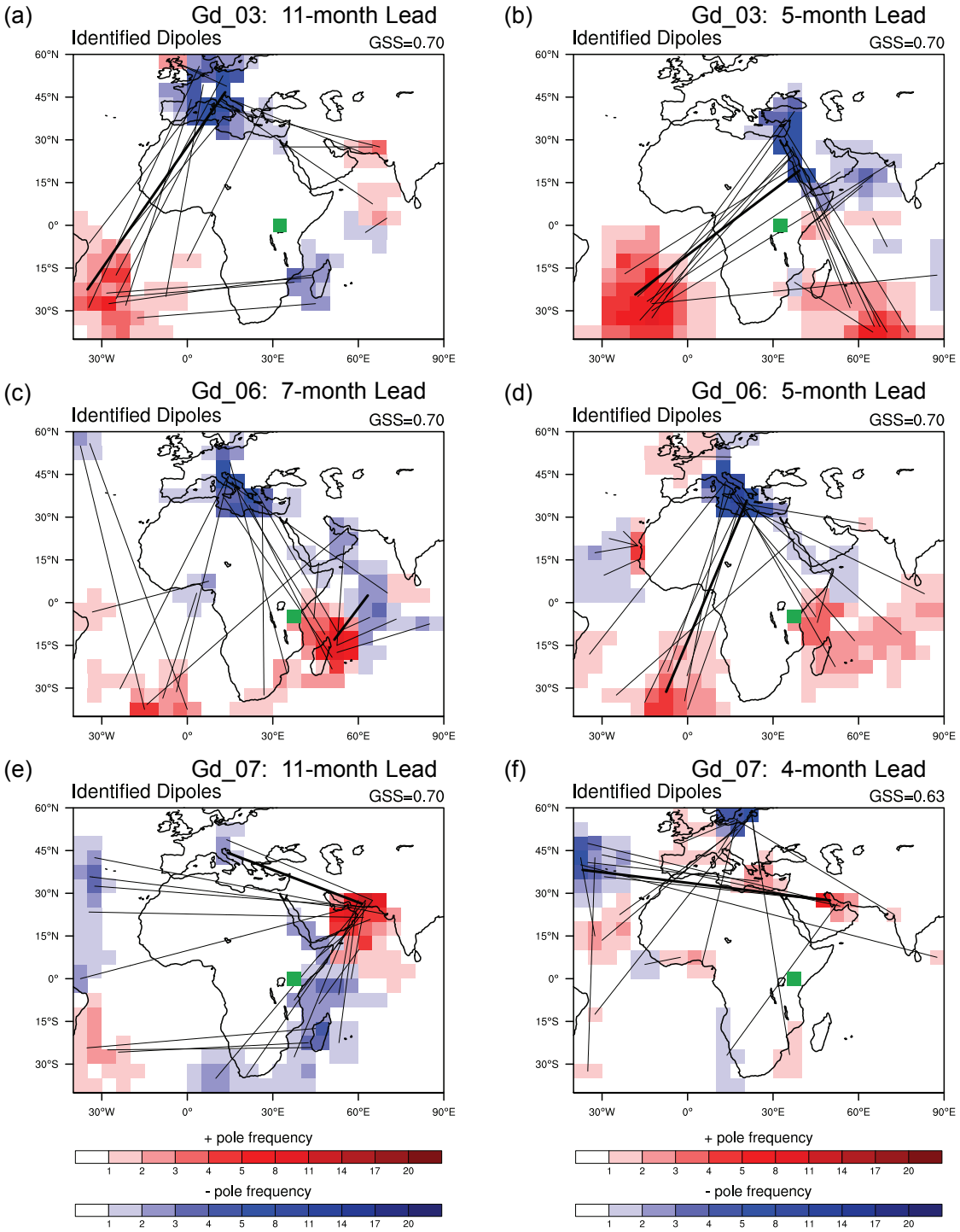



**Figure 5.14:** Corresponding dipoles used to generate the hindcasting results shown in **Figure 5.11**. Grid locations are denoted in green cells. Thicker and thinner connective lines indicate the orientation of the best and remaining dipoles, respectively. The best GSS among all dipoles obtained during the predictor screening process is denoted at the upper-right corner of each figure.

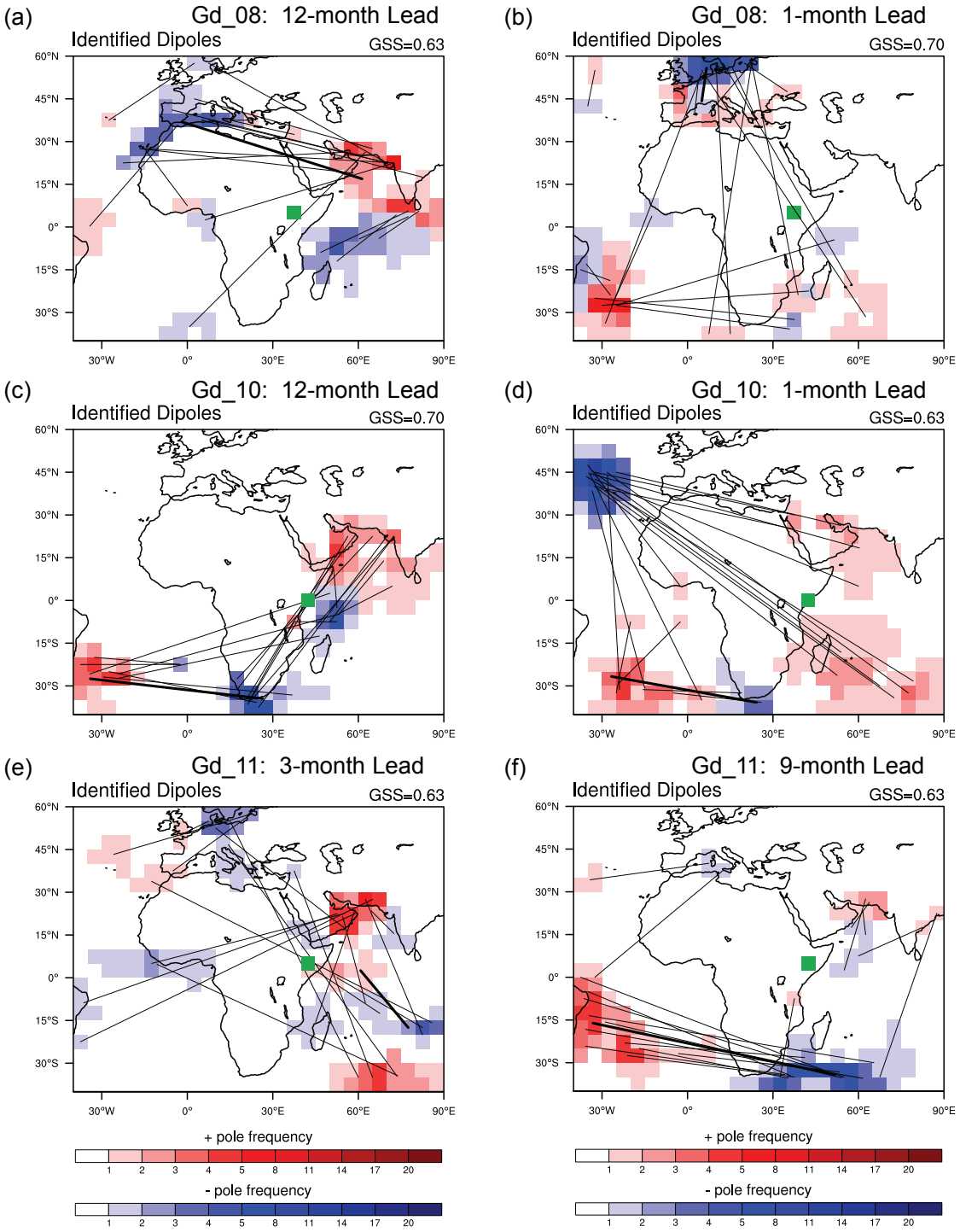



**Figure 5.15:** As in **Figure 5.14**, but for the hindcasting results shown in **Figure 5.12**.

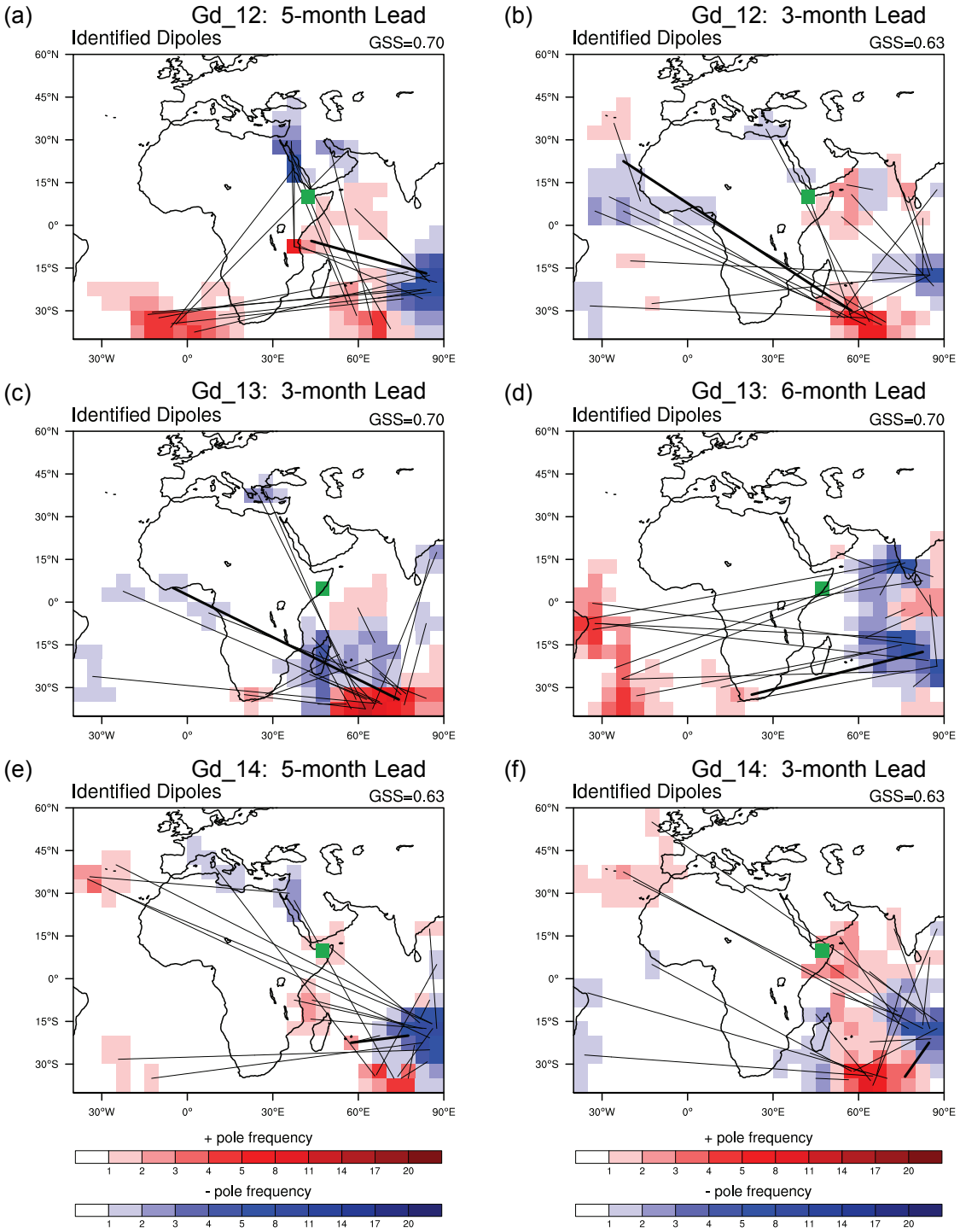



**Figure 5.16:** As in **Figure 5.14**, but for the hindcasting results shown in **Figure 5.13**.

*5.2.3 Tanzania Unimodal Rains*

The third rain type to be analyzed is the T-unimodal rains at the three delineated grids [see **Figure 5.3(c)**]. After applying the algorithm to two SSTA domains (see **Figure 5.4**), it is found that dipoles identified in the long-rains domain can better predict the T-unimodal rains, so the associated forecasting skills are recorded in **Table 5.4**. Among these three grids, the strongest predictability appears at Grid 2 (the mountainous region in northwestern Tanzania) with the highest Re value of 0.767 (at the 9-month lead time) and best CV correlation coefficient of 0.724 (at the 12-month lead time), explaining about 52% of precipitation variance. While the forecasting skill for Grid 5 still remain substantial with Re and $\rho$ values of 0.700 and 0.645, those for Grid 1 become moderate with Re and $\rho$ values of 0.667 and 0.585. However, for all three grids, the optimum forecasting skills consistently occur around long lead times from nine to twelve months.

Thus, hindcasting results at two long lead times (except for Grid 5 where the second lead time is six months) are selected and plotted in **Figure 5.17**. Because the T-unimodal rainy period is as long as five to six months, the average rain rates generally have very little interannual variability. In fact, the observations in the figure reveal that the difference in rain rates between the minimum and maximum years is only two to three mm/day. As a result, the forecasting envelopes for Grids 1 and 2 are relatively flat, but some obvious "turning points" can still be depicted fairly well by the forecasting envelopes at either lead time (e.g., 1991 for Grid 1 and 1993 for Grid 2). Since Grid 5 is located at the coastal region with possibly enhanced coastal convection, the observations and hindcasts seem to exhibit slightly higher variability. However, such flat observations

as well as hindcasts suggest that rainfall events caused by multiple independent mechanisms may exist. Thus, it may be interesting to separate and re-analyze the T-unimodal rains into sub-periods (e.g., early and late seasons).

The corresponding dipoles as shown in **Figure 5.18** present some common patterns at comparable lead times. These patterns include a semi-permanent positive pole at the southwestern Indian Ocean (for all three grids at both lead times) and a negative pole at the subtropical North Atlantic (for Grids 2 and 5 at longer lead times). Along with these common patterns, some other active oceanic areas are also notable for each grid. For example, for Grid 2, positive and negative poles active at the coast of Brazil and the western Mediterranean Sea seem to create tri-pole or even quadri-pole patterns at the 9-month lead time. On the other hand, at the 12-month lead time, the subtropical South Atlantic and the Central Indian Ocean are likely to be active and form another group of dipoles that cross the common dipole patterns. Although these patterns found for Grid 2 have several active oceanic areas, they are more organized in comparison with the patterns for Grids 1 and 5. This is indirectly reflected on the relative forecasting skill of the identified dipoles for these three grids. Such reflection indeed corroborates the physical significance of these patterns, but additional effort is required to discover any underlying mechanisms.

**Table 5.4:** As in **Table 5.2**, but for the Tanzania Unimodal rains.

| Gd_01 | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.404 | 0.498 | **0.667** | -0.671 | 0.612 | 0.371 |
| 2 | 0.389 | **0.585** | 0.600 | -0.559 | 0.533 | 0.399 |
| 3 | 0.399 | 0.516 | 0.567 | -0.615 | 0.644 | 0.379 |
| 4 | 0.422 | 0.357 | 0.533 | -0.660 | 0.651 | 0.365 |
| 5 | 0.429 | 0.376 | 0.567 | -0.537 | 0.559 | 0.378 |
| 6 | 0.424 | 0.354 | 0.467 | -0.636 | 0.787 | **0.344** |
| 7 | 0.405 | 0.450 | 0.533 | -0.752 | 0.779 | 0.368 |
| 8 | 0.378 | 0.530 | 0.600 | -0.427 | 0.661 | 0.363 |
| 9 | 0.388 | 0.505 | **0.667** | -0.616 | 0.808 | 0.386 |
| 10 | **0.373** | 0.486 | **0.667** | -0.664 | 0.701 | 0.364 |
| 11 | 0.386 | 0.449 | **0.667** | -0.487 | **0.469** | 0.396 |
| 12 | 0.384 | 0.501 | 0.600 | **-0.419** | 0.745 | 0.399 |
| **Gd_02** | | | | | | |
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.354 | 0.600 | 0.700 | -0.852 | 1.025 | 0.339 |
| 2 | 0.368 | 0.553 | 0.633 | -0.594 | 0.863 | 0.321 |
| 3 | 0.392 | 0.546 | 0.633 | -0.772 | 1.109 | **0.274** |
| 4 | 0.435 | 0.449 | 0.533 | -0.851 | 1.025 | 0.314 |
| 5 | 0.402 | 0.540 | 0.600 | -0.862 | 1.129 | 0.303 |
| 6 | 0.397 | 0.474 | 0.600 | -0.707 | 1.165 | 0.287 |
| 7 | 0.400 | 0.433 | 0.700 | -0.722 | 0.880 | 0.336 |
| 8 | 0.402 | 0.491 | 0.667 | -0.770 | 1.013 | 0.294 |
| 9 | 0.349 | 0.603 | **0.767** | -0.760 | 0.804 | 0.359 |
| 10 | 0.366 | 0.639 | 0.733 | **-0.551** | 0.712 | 0.368 |
| 11 | 0.357 | 0.567 | 0.733 | -0.716 | 0.621 | 0.376 |
| 12 | **0.331** | **0.724** | 0.667 | -0.641 | **0.351** | 0.375 |

**Table 5.4:** *Continued*.

| Gd_05 | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.512 | 0.580 | 0.533 | -0.448 | 1.152 | 0.395 |
| 2 | 0.564 | 0.466 | 0.467 | -0.682 | 1.284 | 0.358 |
| 3 | 0.576 | 0.532 | 0.500 | -0.426 | 0.882 | 0.369 |
| 4 | 0.606 | 0.354 | 0.433 | -0.636 | 1.363 | **0.317** |
| 5 | 0.593 | 0.366 | 0.467 | -0.594 | 1.206 | 0.336 |
| 6 | 0.524 | 0.556 | 0.600 | -0.466 | 1.027 | 0.396 |
| 7 | 0.541 | 0.538 | 0.533 | -0.527 | 0.696 | 0.419 |
| 8 | 0.554 | 0.561 | 0.500 | -0.610 | 0.712 | 0.412 |
| 9 | 0.561 | 0.472 | 0.433 | -0.736 | 0.943 | 0.348 |
| 10 | 0.522 | 0.614 | 0.600 | -0.497 | 0.989 | 0.410 |
| 11 | 0.505 | **0.645** | **0.700** | **-0.181** | **0.629** | 0.458 |
| 12 | **0.494** | 0.620 | **0.700** | -0.501 | 0.738 | 0.403 |

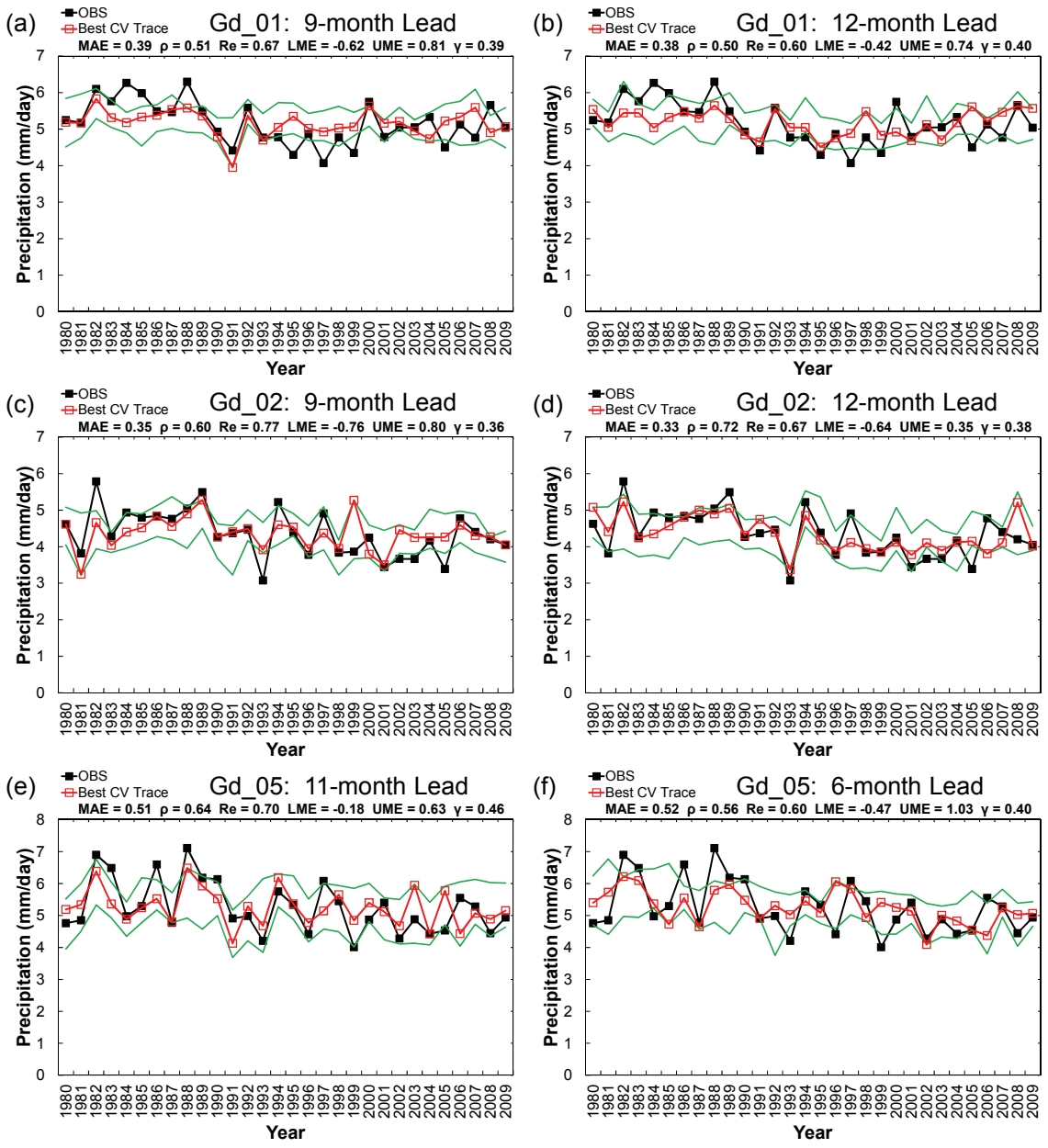



**Figure 5.17:** Optimal hindcasting results at two selected lead times for the T-unimodal rains at Grids 1, 2, and 5. The best CV trace is shown in red, and the green envelope is constructed using the maximum and minimum hindcasting values.

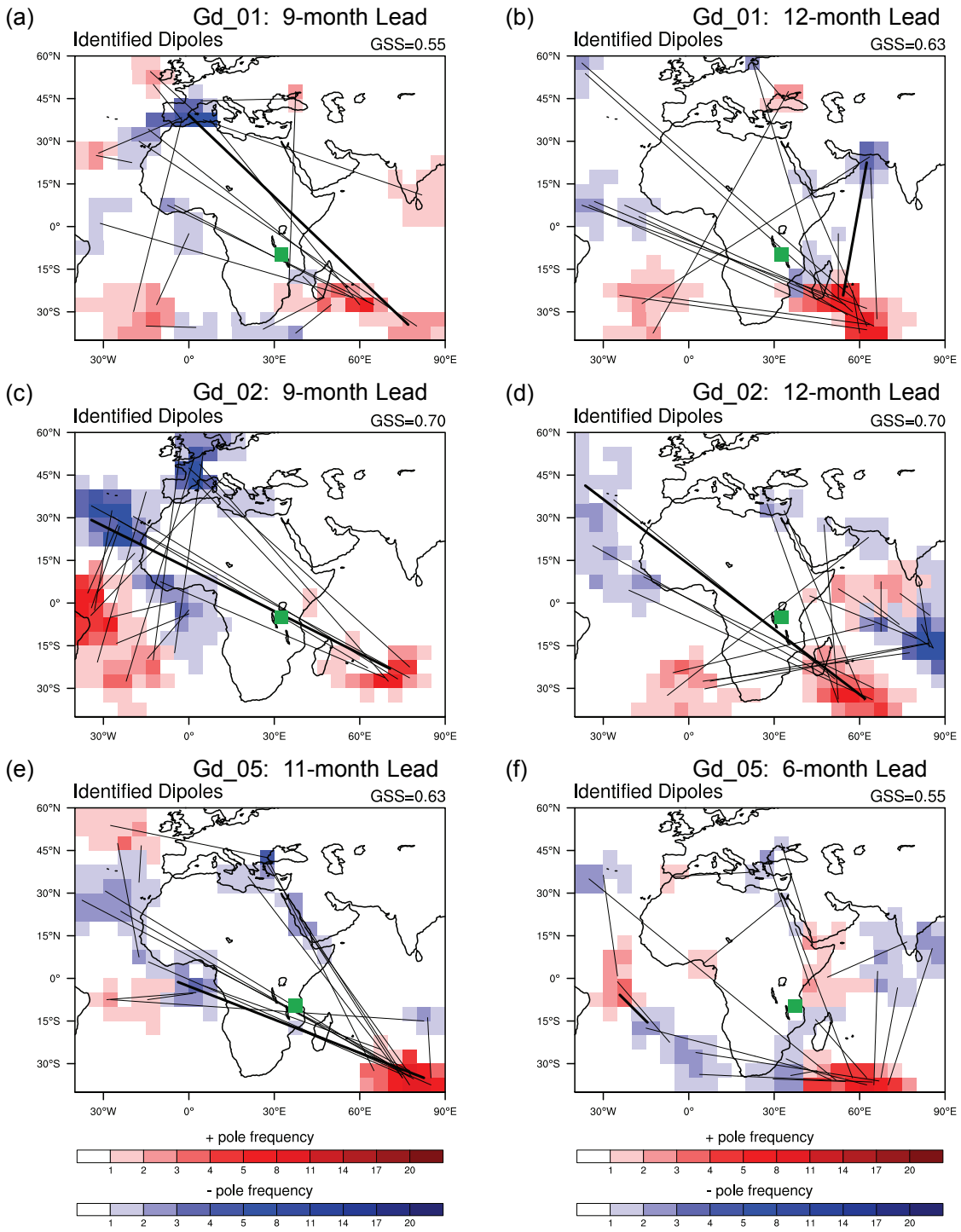



**Figure 5.18:** Corresponding dipoles used to generate the hindcasting results shown in **Figure 5.17**. Grid locations are denoted in green cells. Thicker and thinner connective lines indicate the orientation of the best and remaining dipoles, respectively. The best GSS among all dipoles obtained during the predictor screening process is denoted at the upper-right corner of each figure.

*5.2.4 Southern Sudan and Kiremt Rains*

The last two rain types examined in this study are the southern Sudan and the Kiremt rains at the three delineated grids [see **Figure 5.3(d)**]. When the long-rains domain is adopted, the algorithm is able to produce better forecasting skills for the southern Sudan rains as summarized in **Table 5.5**. The highest Re value is 0.667, and the best CV correlation coefficient is 0.562, only accounting for about 30% of the precipitation variance. Such moderate skill is possibly attributed to the very little interannual variability of the southern Sudan rains. Similar to the T-unimodal rains, the southern Sudan rainy period can be as long as seven months with insignificant intra-seasonal variations. Therefore, the seven-month average only generates a maximum interannual difference in rain rates around 2.5 mm/day, as shown by the 30-year observations in **Figure 5.19**.

Because the wet and dry events of this rain type may not be strong enough to reflect significant climate oscillations, the dipole algorithm seems to have difficulty finding prominent predictors in the defined SSTA field. Apart from a tripole-like pattern composed of a negative pole at the eastern Mediterranean Sea (even up to the Black Sea) and two positive poles at the South Atlantic and South Indian Oceans [**Figure 5.20(a)**], the remaining patterns appear to be more disorganized than those found for other rain types. This may also explain why the forecasting traces and envelopes are relatively flat.

**Table 5.5:** As in **Table 5.2**, but for the southern Sudan rains.

| Gd_04 | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.374 | 0.303 | 0.433 | -1.126 | 0.924 | **0.263** |
| 2 | 0.360 | 0.448 | 0.467 | -1.138 | 0.645 | 0.279 |
| 3 | 0.343 | 0.492 | 0.633 | -0.972 | 0.603 | 0.297 |
| 4 | 0.328 | 0.513 | 0.533 | -0.870 | 0.774 | 0.292 |
| 5 | 0.356 | 0.360 | 0.533 | -0.887 | 0.872 | 0.291 |
| 6 | 0.341 | 0.509 | 0.533 | **-0.741** | 0.779 | 0.273 |
| 7 | 0.351 | 0.454 | 0.467 | -0.743 | 0.866 | 0.290 |
| 8 | 0.344 | 0.412 | 0.467 | -0.841 | 0.725 | 0.296 |
| 9 | 0.332 | 0.521 | 0.567 | -0.811 | 0.571 | 0.290 |
| 10 | **0.324** | **0.562** | **0.667** | -0.763 | **0.343** | 0.308 |
| 11 | 0.347 | 0.387 | 0.567 | -0.891 | 0.593 | 0.306 |
| 12 | 0.342 | 0.408 | 0.567 | -0.907 | 0.894 | 0.311 |

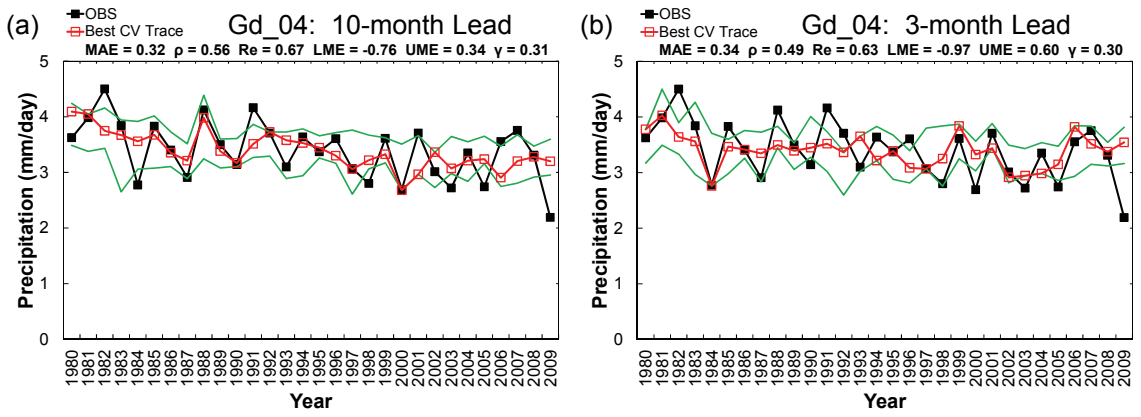



**Figure 5.19:** Optimal hindcasting results at two selected lead times for the southern Sudan rains at Grid 4. The best CV trace is shown in red, and the green envelope is constructed using the maximum and minimum hindcasting values.

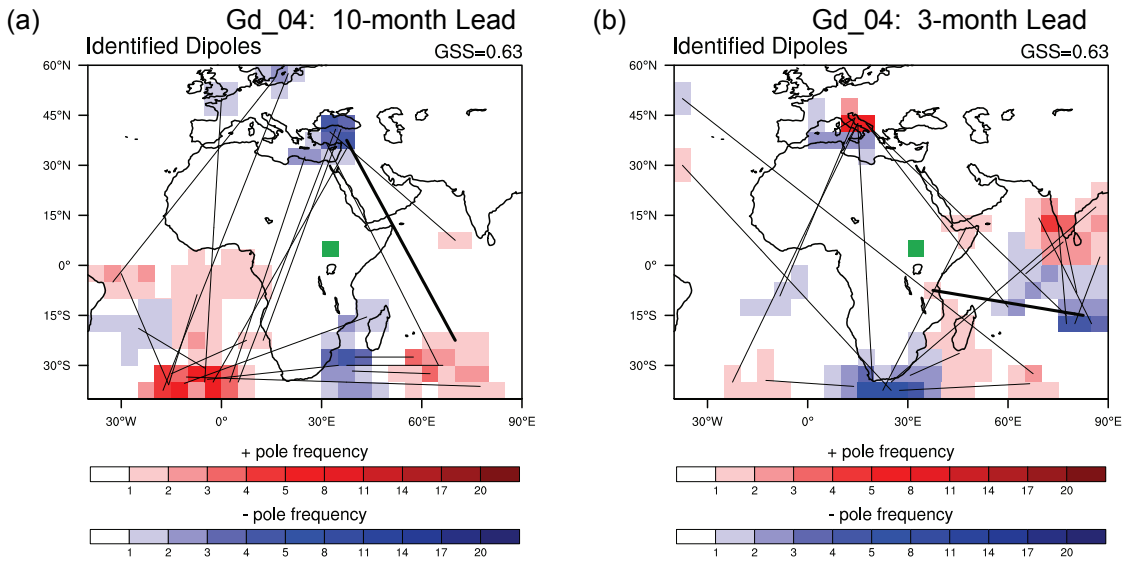



**Figure 5.20:** Corresponding dipoles used to generate the hindcasting results shown in **Figure 5.19**. Grid locations are denoted in green cells. Thicker and thinner connective lines indicate the orientation of the best and remaining dipoles, respectively. The best GSS among all dipoles obtained during the predictor screening process is denoted at the upper-right corner of each figure.

Regarding the Kiremt rains, although slightly better forecasting skill is produced when the long-rains SSTA domain is adopted, the identified dipoles indicate that the dominant signal still emanates from the Indian Ocean (**Figure 5.21**). The slight improvement in forecasting skills generally results from the increase in spatial randomness from the Atlantic Ocean. To incorporate with the most plausible physics, the associated forecasting skill obtained from applying the short-rains domain are chosen and presented in **Table 5.6**. The skill for both grids is comparable with the highest Re value of 0.667 and best CV correlation coefficients of 0.609 (Grid 9) and 0.664 (Grid 12) at median to long lead times. Since the onset of the Kiremt rains at Grid 9 (June) is one month earlier

than that at Grid 12 (July), the relatively optimal lead times (in terms of Re values) actually indicate the same season (DJF).

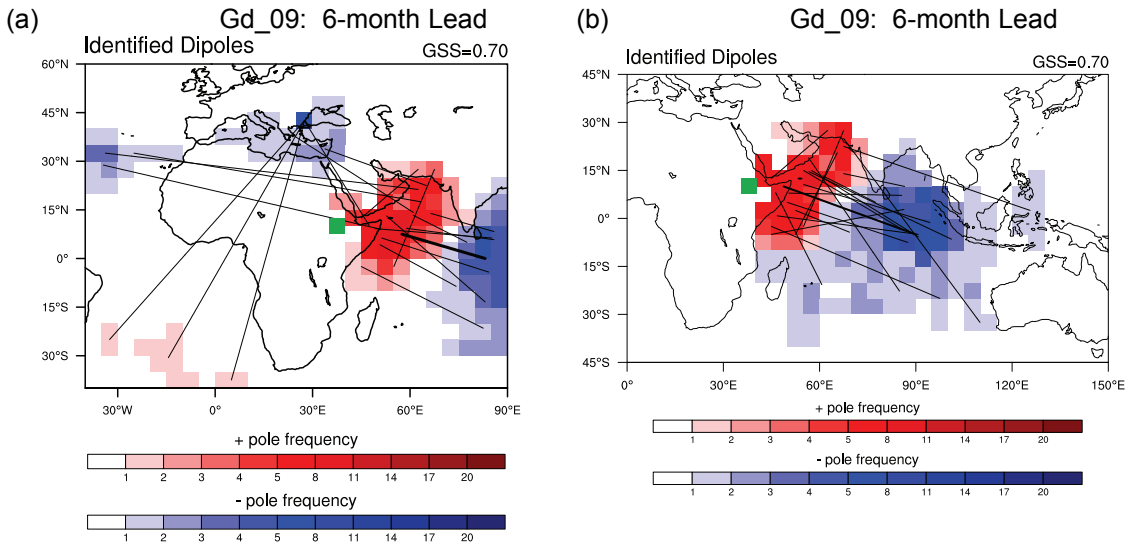



**Figure 5.21:** Comparison of the identified dipoles between (a) using the long-rains domain and (b) using the short-rains domain for the Kiremt rains at Grid 9.

To explore the forecasting skills and detailed correspondence between the forecasting envelopes and observations, the associated results at two different lead times are again selected and plotted in **Figure 5.22**. From the figure, the observations of the Kiremt rains at these two grids show quite different characteristics. Because Grid 9 is located at the mountainous region of Ethiopia with possible orographic lifting effects, it tends to receive more rain than Grid 12. Besides, the wet and dry years at Grid 9 do not coincide well with those at Grid 12. For instance, 1990 and 1993 are wet at Grid 9, while the same years are dry at Grid 12. In light of these contrasts, the way in which the forecasting envelopes respond to the variations in the observations at the two grids is also different.

Although the forecasting envelopes can remarkably contain or respond to most of the observations at both grids, some low- and high-value discrepancies are more noticeable for either Grid 9 or 12 (e.g., 1982 and 2009 for Grid 9; 1981 and 1994 for Grid 12).

The identified dipoles used to generate the hindcasting results at two different lead times are shown in **Figure 5.23**. At a median lead time, the zonal-oriented dipoles identified for Grid 9, composed of concentrated positive and negative poles at East Africa's coast to the Arabian Sea and the Central to East Indian Ocean, seem to be more organized than those for Grid 12. Besides, negative pole areas are likely to be less important than another active positive pole areas in the vicinity of Indonesia for the latter grid. On the other hand, at a long lead time, some clockwise-like migration of both positive (shift to the North Indian Ocean to the Bay of Bengal) and negative poles (shift to the subtropical South Indian Ocean) can be observed for Grid 9, and the consequential meridional-oriented dipoles for both grids become more alike. All these SSTA dipoles presumably illustrate their connections with the formation and intensity of the tradewinds and monsoon systems over the Indian Ocean prevailing in early seasons, which may subsequently influence the moisture supply to the following Kiremt rainy season.

**Table 5.6:** As in **Table 5.2**, but for the Kiremt rains in Ethiopia.

| Gd_09 | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.479 | 0.509 | 0.467 | -1.271 | 1.197 | 0.269 |
| 2 | 0.577 | 0.131 | 0.267 | -1.756 | 1.293 | - |
| 3 | 0.542 | 0.181 | 0.367 | -1.611 | 1.227 | **0.210** |
| 4 | 0.522 | 0.322 | 0.500 | -1.638 | 1.201 | 0.243 |
| 5 | 0.456 | 0.543 | 0.567 | -1.558 | 0.665 | 0.288 |
| 6 | **0.415** | **0.609** | **0.667** | -1.476 | **0.599** | 0.300 |
| 7 | 0.450 | 0.543 | 0.600 | -1.041 | 0.774 | 0.292 |
| 8 | 0.464 | 0.496 | 0.567 | -1.147 | 0.767 | 0.308 |
| 9 | 0.464 | 0.536 | 0.600 | -1.329 | 0.657 | 0.335 |
| 10 | 0.473 | 0.502 | 0.500 | -1.263 | 0.816 | 0.311 |
| 11 | 0.477 | 0.585 | 0.467 | **-0.992** | 0.825 | 0.293 |
| 12 | 0.487 | 0.589 | 0.567 | -1.020 | 0.973 | 0.322 |
| **Gd_12** | | | | | | |
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.410 | 0.463 | 0.433 | -0.619 | 1.341 | 0.266 |
| 2 | 0.407 | 0.206 | 0.567 | -0.842 | 1.408 | 0.246 |
| 3 | 0.415 | 0.329 | 0.400 | -0.686 | 1.420 | **0.204** |
| 4 | 0.420 | 0.293 | 0.433 | -0.573 | 1.576 | 0.211 |
| 5 | 0.364 | 0.499 | 0.633 | -0.603 | **1.105** | 0.269 |
| 6 | 0.361 | 0.615 | 0.600 | **-0.572** | 1.176 | 0.270 |
| 7 | 0.339 | 0.504 | **0.667** | -0.657 | 1.385 | 0.271 |
| 8 | 0.348 | **0.664** | 0.600 | -0.630 | 1.257 | 0.278 |
| 9 | 0.357 | 0.619 | 0.600 | -0.594 | 1.334 | 0.293 |
| 10 | 0.337 | 0.618 | **0.667** | -0.721 | 1.275 | 0.285 |
| 11 | **0.327** | 0.618 | 0.567 | -0.580 | 1.403 | 0.290 |
| 12 | 0.349 | 0.602 | 0.567 | -0.683 | 1.276 | 0.263 |

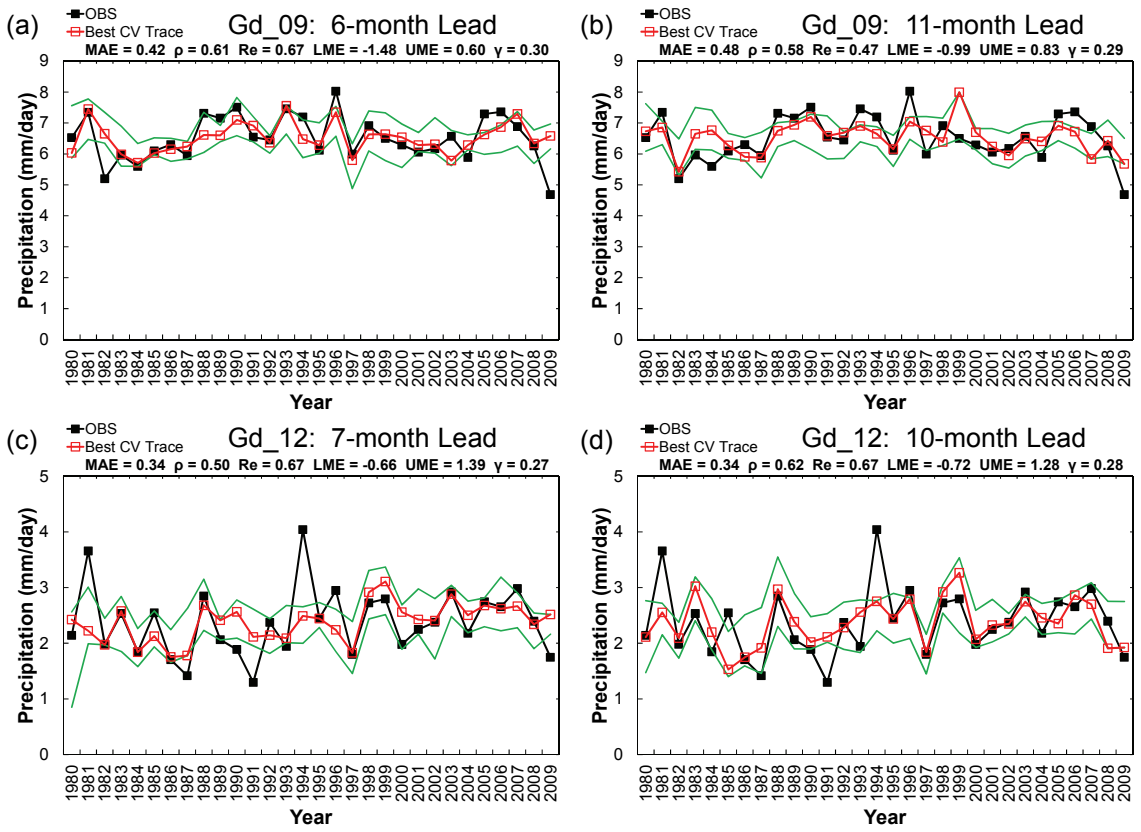



**Figure 5.22:** Optimal hindcasting results at two selected lead times for the Kiremt rains at Grids 9 and 12. The best CV trace is shown in red, and the green envelope is constructed using the maximum and minimum hindcasting values.

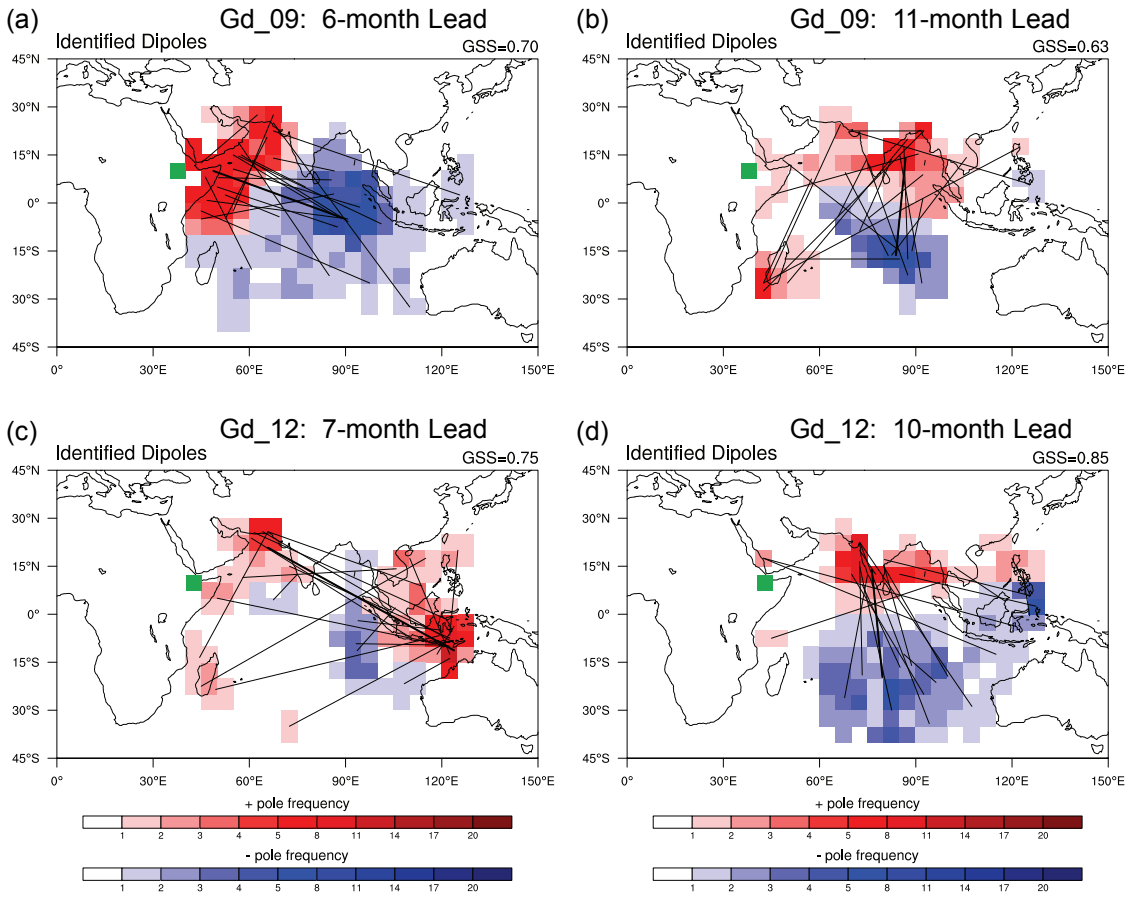



**Figure 5.23:** Corresponding dipoles used to generate the hindcasting results shown in **Figure 5.22**. Grid locations are denoted in green cells. Thicker and thinner connective lines indicate the orientation of the best and remaining dipoles, respectively. The best GSS among all dipoles obtained during the predictor screening process is denoted at the upper-right corner of each figure.

## 5.3 Sensitivity Analysis of SSTA Domains

Even though most of the hindcasting results and identified dipoles thus far presented are significant, further improvement in forecasting skills may be required (especially for the short rains, which nearly have no forecasting implication). Such improvement can possibly be accomplished by enlarging the defined SSTA domains to create more spatial predictors. In addition, many studies debated that ENSO may play a dominated role in modulating the rainy period in East Africa (e.g., Janowiak, 1988; Ogallo, 1988; Mutai et al., 1998; Indeje et al., 2000; Gissila et al., 2004; Segele and Lamb, 2005; Block and Rajagopalan, 2007; Korecha and Barnston, 2007; Segele et al., 2009; Diro et al., 2011a), which can also be verified by employing the dipole algorithm over the Pacific Ocean. Therefore, to answer these questions, both the short- and long-rains domains are enlarged eastward to cover the eastern tropical Pacific (the key ENSO region). Such enlarged domains are then applied to the dipole algorithm to rerun for selected grids that previously exhibited more significant hindcasting results and/or prominent dipole patterns. These grids include Grids 3 and 11 for both the short and long rains and Grid 9 for the Kiremt rains.

### *5.3.1 Short Rains at Grids 3 and 11*

The updated forecasting skills after applying a larger SSTA domain to the dipole algorithm for the short rains at Grids 3 and 11 are recorded in **Table 5.7**. The highest skills for Grid 3 at near-concurrent lead times increase substantially to highest Re value of 0.833 (from 0.733) and best CV correlation coefficient of 0.789 (from 0.714). On the other hand, the highest skill for Grid 11 seems to only increase moderately with highest

Re value remaining as before (0.900) and best CV correlation coefficient of 0.720 (from 0.707). More importantly, the skill for both grids at longer lead times ($\geq 3$ months) now becomes more substantial and applicable to forecasting practices.

To observe the variations of the hindcasting results and the corresponding dipoles after using a larger SSTA domain, more significant results at two lead times (with the same logic as previously described) for Grids 3 and 11 are selected and plotted in **Figures 5.24** and **5.25**, respectively. For Grid 3 at the 1-month lead time, while the major positive poles remain active at the south coast of Australia (with some extension to the Central Indian Ocean), the negative poles migrate northeastward to the western Pacific Ocean [**Figures 5.24(b)** and **5.24(d)**]. Accompanied with some other patterns newly identified over the Pacific Ocean, these dipoles create a forecasting envelope that better corresponds with the observations in certain periods (e.g., 2007–2009). At the 9-month lead time, the originally disorganized dipoles that only generate moderate forecasting skills are largely replaced by more pronounced dipoles composed of vast negative poles over the Indian Ocean, a concentrated positive pole at the South Pacific, and even some ENSO signals after introducing a larger domain [**Figures 5.24(f)** and **5.24(h)**].

For Grid 11 at the 1-month lead time, although parts of the positive and negative poles slightly shift or extend eastward, the general patterns over the Indian Ocean remain and seem to be more pronounced than scattered signals from the Pacific Ocean [**Figures 5.25(b)** and **5.25(d)**] thereby generating very similar forecasting skills with only marginal improvement. At the 3-month lead time, again the dipole patterns over the Indian Ocean

remain nearly unchanged, but a more concentrated positive pole related to ENSO is identified [**Figures 5.25(f)** and **5.25(h)**], which is able to create more variability for the forecasting envelope so that the Re value is improved. Notably, the ENSO signals seem to experience a phase transition from long to short lead times. Whether this transition is substantial will further be verified through composite analysis carried out in Section 5.5.

Overall, this analysis indicates that including the Pacific Ocean indeed exhibits contributions to the forecasting of the short rains, and the ENSO signals do exist at certain lead times. However, a conjunctive use of SSTA information in both the Indian and Pacific Oceans is still essential to ensure optimum forecasting skill for the short rains, as demonstrated by this analysis.

**Table 5.7:** Updated forecasting skills for the short rains at Grids 3 and 11 after using a larger SSTA domain.

| Gd_03 (Larger Domain) | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | **0.469** | **0.789** | **0.833** | **-0.382** | 1.449 | 0.399 |
| 2 | 0.472 | 0.722 | **0.833** | -0.755 | 1.152 | 0.385 |
| 3 | 0.520 | 0.697 | 0.767 | -0.695 | 1.683 | 0.397 |
| 4 | 0.559 | 0.661 | 0.767 | -1.295 | 1.896 | 0.351 |
| 5 | 0.558 | 0.670 | 0.733 | -1.094 | 1.466 | 0.351 |
| 6 | 0.548 | 0.693 | 0.767 | -1.128 | 1.483 | 0.356 |
| 7 | 0.612 | 0.527 | 0.633 | -0.902 | 1.397 | 0.346 |
| 8 | 0.562 | 0.598 | 0.733 | -1.031 | 1.355 | 0.361 |
| 9 | 0.518 | 0.740 | 0.767 | -1.139 | **0.911** | 0.362 |
| 10 | 0.535 | 0.710 | 0.767 | -0.847 | 1.009 | 0.387 |
| 11 | 0.625 | 0.564 | 0.633 | -1.190 | 1.744 | **0.337** |
| 12 | 0.657 | 0.462 | 0.567 | -1.490 | 1.612 | 0.380 |
| **Gd_11 (Larger Domain)** | | | | | | |
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | **0.461** | 0.696 | **0.900** | **-0.245** | **2.030** | 0.342 |
| 2 | 0.514 | **0.720** | 0.833 | -0.361 | 2.496 | 0.306 |
| 3 | 0.524 | 0.566 | 0.800 | -0.498 | 2.584 | 0.321 |
| 4 | 0.627 | 0.410 | 0.633 | -0.808 | 3.179 | 0.244 |
| 5 | 0.587 | 0.481 | 0.633 | -0.619 | 3.472 | 0.214 |
| 6 | 0.554 | 0.497 | 0.667 | -0.582 | 3.310 | 0.236 |
| 7 | 0.559 | 0.478 | 0.567 | -0.552 | 3.422 | 0.233 |
| 8 | 0.570 | 0.450 | 0.667 | -0.628 | 3.528 | 0.227 |
| 9 | 0.630 | 0.244 | 0.567 | -0.912 | 3.548 | 0.212 |
| 10 | 0.634 | 0.315 | 0.633 | -0.828 | 3.557 | **0.203** |
| 11 | 0.649 | 0.311 | 0.600 | -0.642 | 3.813 | 0.204 |
| 12 | 0.613 | 0.404 | 0.633 | -0.855 | 3.542 | 0.214 |

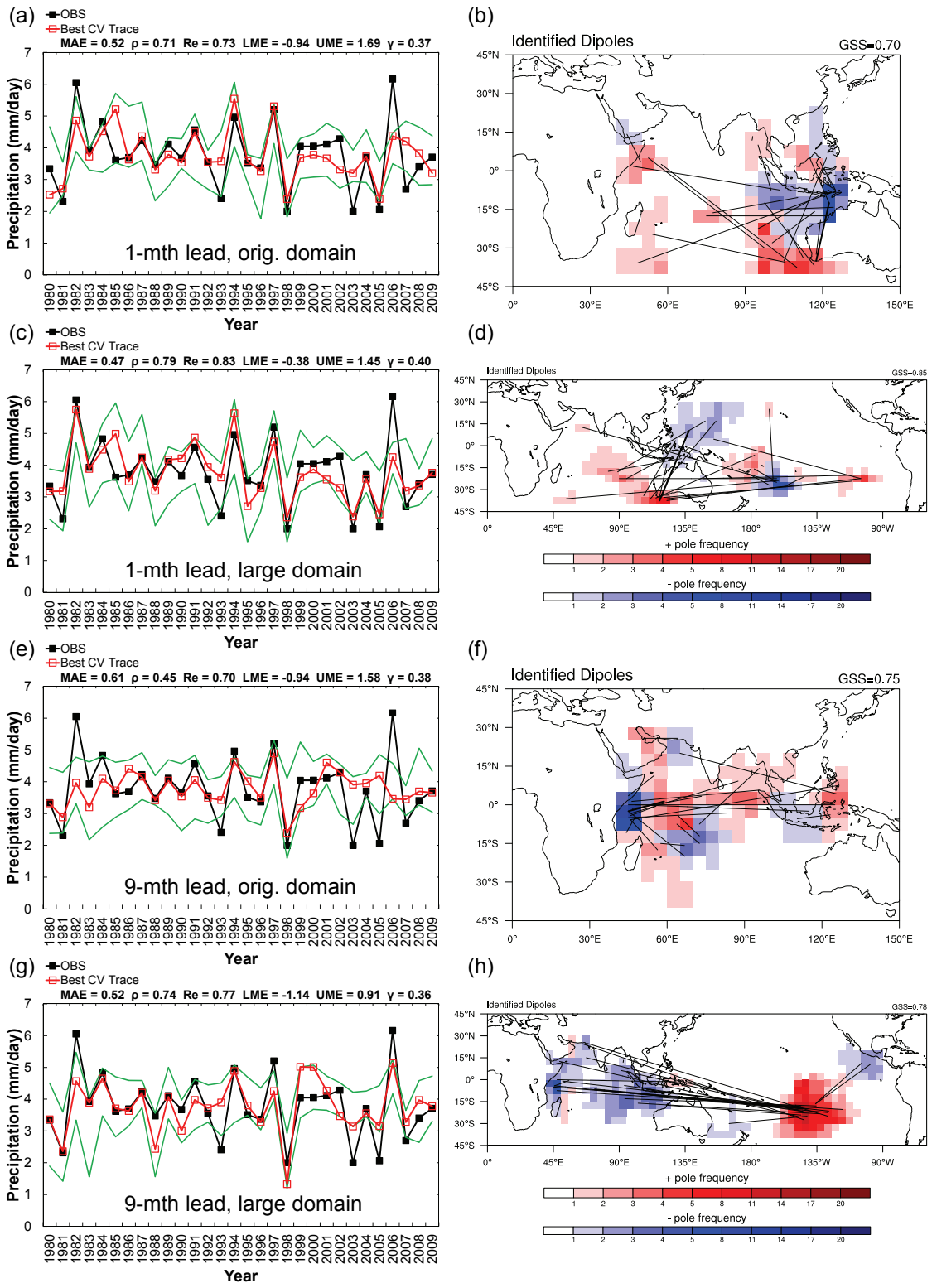



**Figure 5.24:** Variations of the hindcasting results and the corresponding dipoles for the short rains at Grid 3 after using a larger SSTA domain; results at 1- and 9-month lead times are shown.

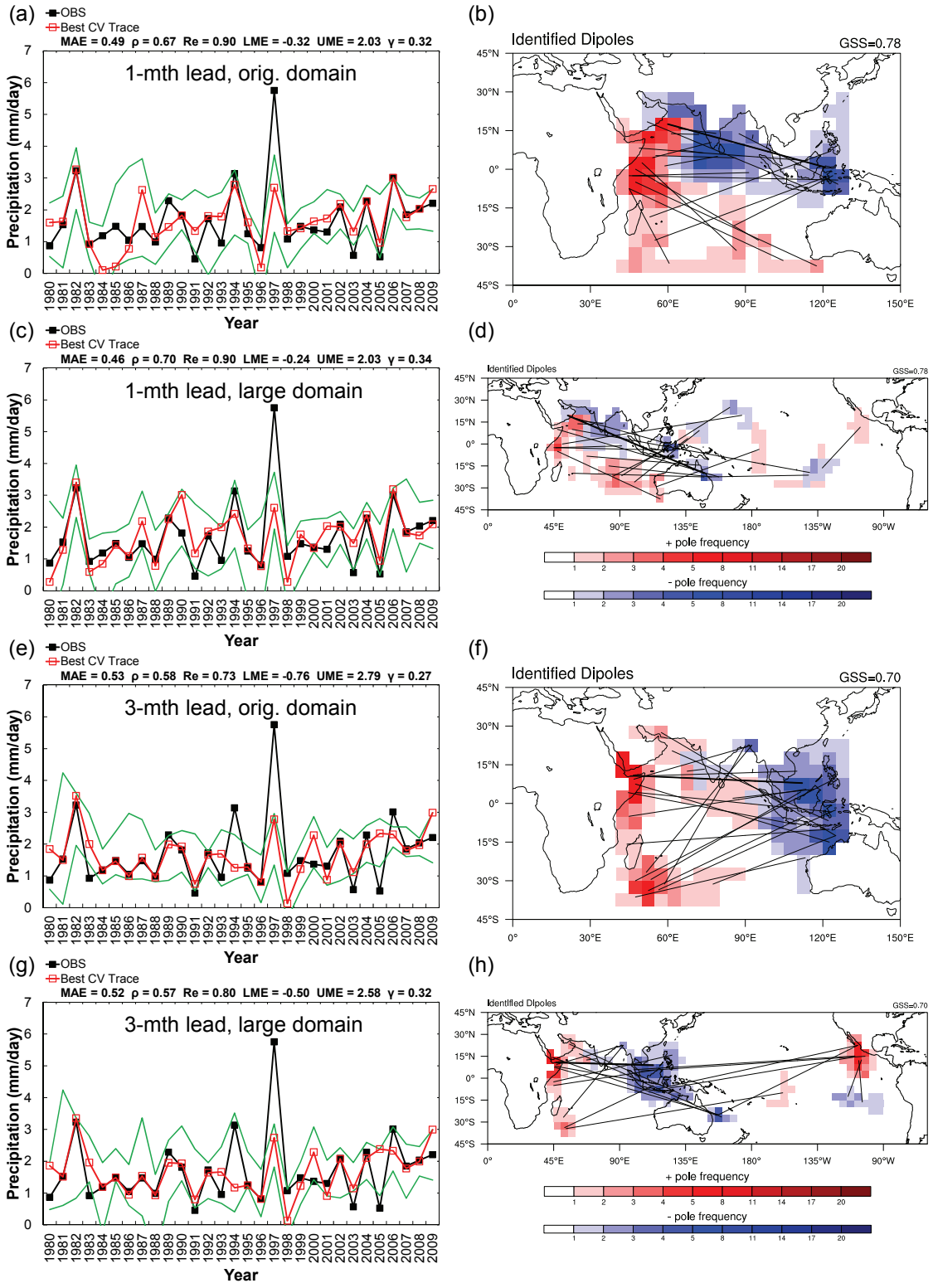



**Figure 5.25:** As in **Figure 5.24**, but for the short rains at Grid 11; results at 1- and 3-month lead times are shown.

*5.3.2 Long Rains at Grids 3 and 11*

The updated forecasting skill under a larger SSTA domain for the long rains at Grids 3 and 11 (**Table 5.8**) reveals similar positive effects. The highest skills for Grid 3 at long lead times increase substantially with highest Re value of 0.867 (from 0.800) and best CV correlation coefficient of 0.774 (from 0.758). The skill improvement for Grid 11 is even more significant: The highest Re value increases to 0.833 (from 0.667), and the best CV correlation coefficient increases to 0.763 (from 0.614). The relative optimum forecasting skill emerges from 6- to 9-month lead times when only moderate to low skill was found previously, implying that some newly identified signals from the Pacific Ocean are more dominant during these periods.

**Figure 5.26** shows how the hindcasting results and associated dipoles vary at two selected lead times for Grid 3 after using the larger SSTA domain. At the 9-month lead time, while the dipole components at the Mediterranean to Black Seas and the South Atlantic remain active, some positive and negative poles formerly located in the Indian Ocean are eliminated from the newly identified dipoles. The remaining patterns connected with positive poles over the western Pacific Ocean and some ENSO-related signals [**Figure 5.26(d)**] seem to improve the reliability of the forecasting envelope. Likewise, at the 5-month lead time, the original patterns at the Indian Ocean are largely replaced by new dipole combinations over other oceanic areas, such as pronounced dipoles over the western Pacific Ocean and negative poles over the North Atlantic [**Figure 5.26(h)**].

While the algorithm still retains some of the major dipoles (at the west side of the SSTA domain) for Grid 3, it generally excludes most of the original dipoles for Grid 11 after the SSTA domain is enlarged. At the 6-month lead time, except for the negative pole in the Indian Ocean, almost all scattered dipoles are replaced by more organized dipoles in the western Pacific Ocean and some other patterns at the South Pacific and the south coast of Australia [**Figure 5.27(d)**]. This new set of dipole predictors is able to adjust several obvious discrepancies between the observations and ensemble forecasts (e.g., in 1987, 1994, and 2000). At the 9-month lead time, although the original dipoles appear to be organized, they are not as significant as new signals from Indonesia and the ENSO region [**Figure 5.27(h)**].

Including the Pacific Ocean is again shown to be beneficial for forecasting of the long rains. ENSO signals appear at the long lead time but with a different sign for each grid. This may not only explain the contradictory results presented by previous studies but also demonstrate the necessity of an efficient predictor screening algorithm.

**Table 5.8:** As in **Table 5.7**, but for the long rains at Grids 3 and 11.

| Gd_03 (Larger Domain) | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.451 | 0.666 | 0.767 | -0.856 | 1.457 | 0.356 |
| 2 | 0.472 | 0.621 | 0.800 | -0.539 | 1.023 | **0.346** |
| 3 | 0.456 | 0.699 | 0.767 | **-0.383** | 1.064 | 0.360 |
| 4 | 0.428 | 0.699 | 0.733 | -0.419 | 1.138 | 0.359 |
| 5 | 0.442 | 0.718 | 0.833 | -0.766 | 0.829 | 0.374 |
| 6 | 0.452 | 0.704 | 0.767 | -0.623 | 0.782 | 0.385 |
| 7 | 0.490 | 0.669 | 0.800 | -0.727 | 1.072 | 0.383 |
| 8 | 0.479 | 0.688 | 0.833 | -0.838 | **0.422** | 0.384 |
| 9 | 0.419 | **0.774** | **0.867** | -0.696 | 0.657 | 0.396 |
| 10 | 0.427 | 0.729 | **0.867** | -0.646 | 0.841 | 0.384 |
| 11 | 0.453 | 0.623 | 0.800 | -0.753 | 0.647 | 0.350 |
| 12 | **0.396** | 0.734 | 0.800 | -0.706 | 0.526 | 0.351 |
| **Gd_11 (Larger Domain)** | | | | | | |
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.414 | 0.524 | 0.667 | -0.546 | 0.916 | 0.356 |
| 2 | 0.363 | 0.510 | 0.667 | -0.297 | 0.677 | 0.396 |
| 3 | 0.379 | 0.593 | 0.733 | -0.145 | 0.781 | 0.389 |
| 4 | 0.363 | 0.642 | 0.667 | -0.633 | 0.853 | **0.343** |
| 5 | 0.344 | 0.581 | 0.733 | -0.145 | 0.579 | 0.360 |
| 6 | **0.318** | 0.728 | **0.833** | **-0.070** | 0.454 | 0.415 |
| 7 | 0.339 | 0.694 | **0.833** | -0.201 | 0.665 | 0.375 |
| 8 | 0.333 | 0.712 | 0.667 | -0.381 | 0.997 | 0.354 |
| 9 | 0.365 | 0.571 | **0.833** | -0.311 | 0.746 | 0.375 |
| 10 | 0.320 | **0.763** | 0.800 | -0.347 | 0.371 | 0.351 |
| 11 | 0.347 | 0.651 | 0.800 | -0.191 | 0.805 | 0.419 |
| 12 | 0.334 | 0.709 | 0.700 | -0.143 | **0.226** | 0.374 |

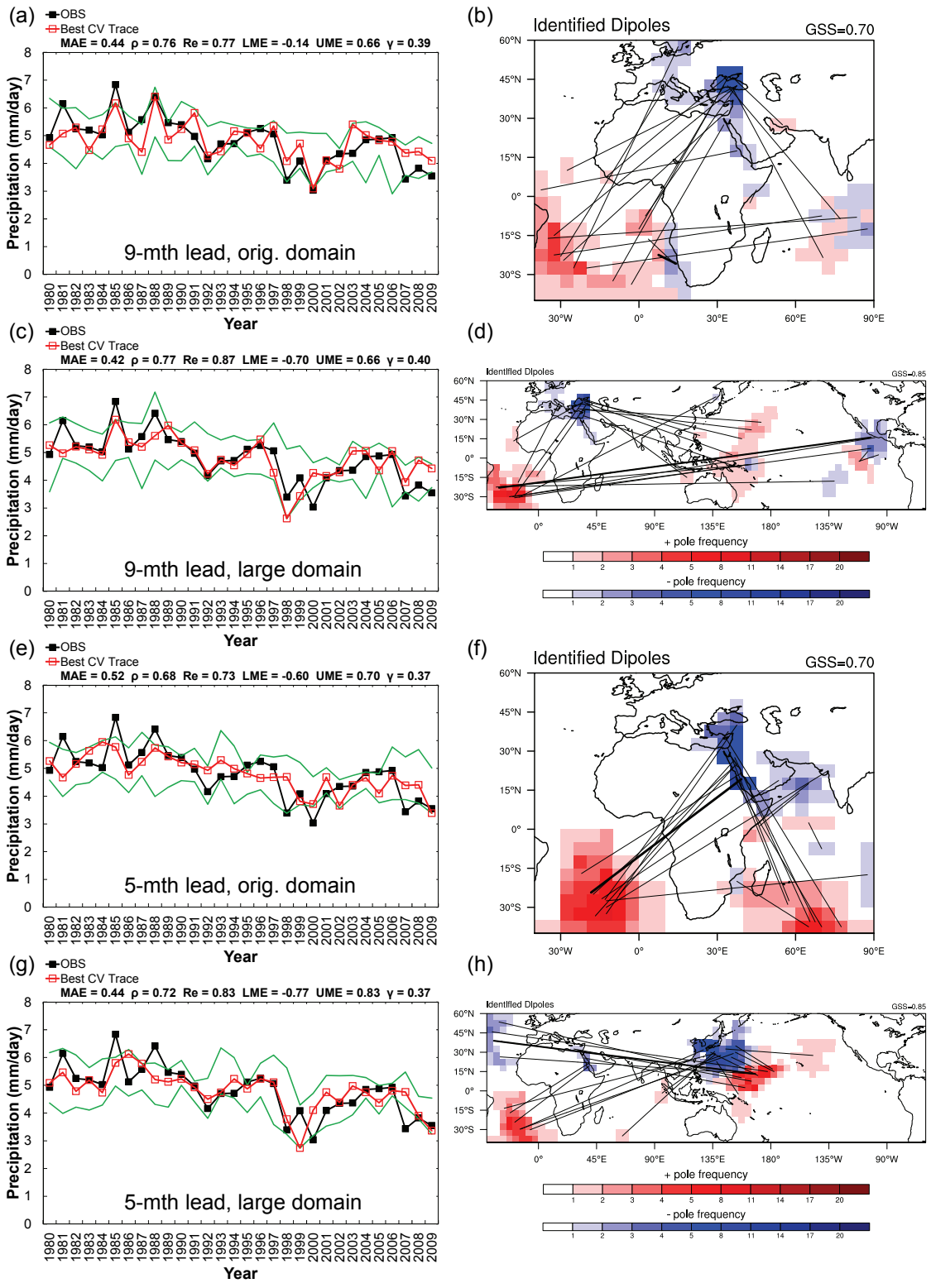



**Figure 5.26:** As in **Figure 5.24**, but for the long rains at Grid 3; results at 9- and 5-month lead times are shown.

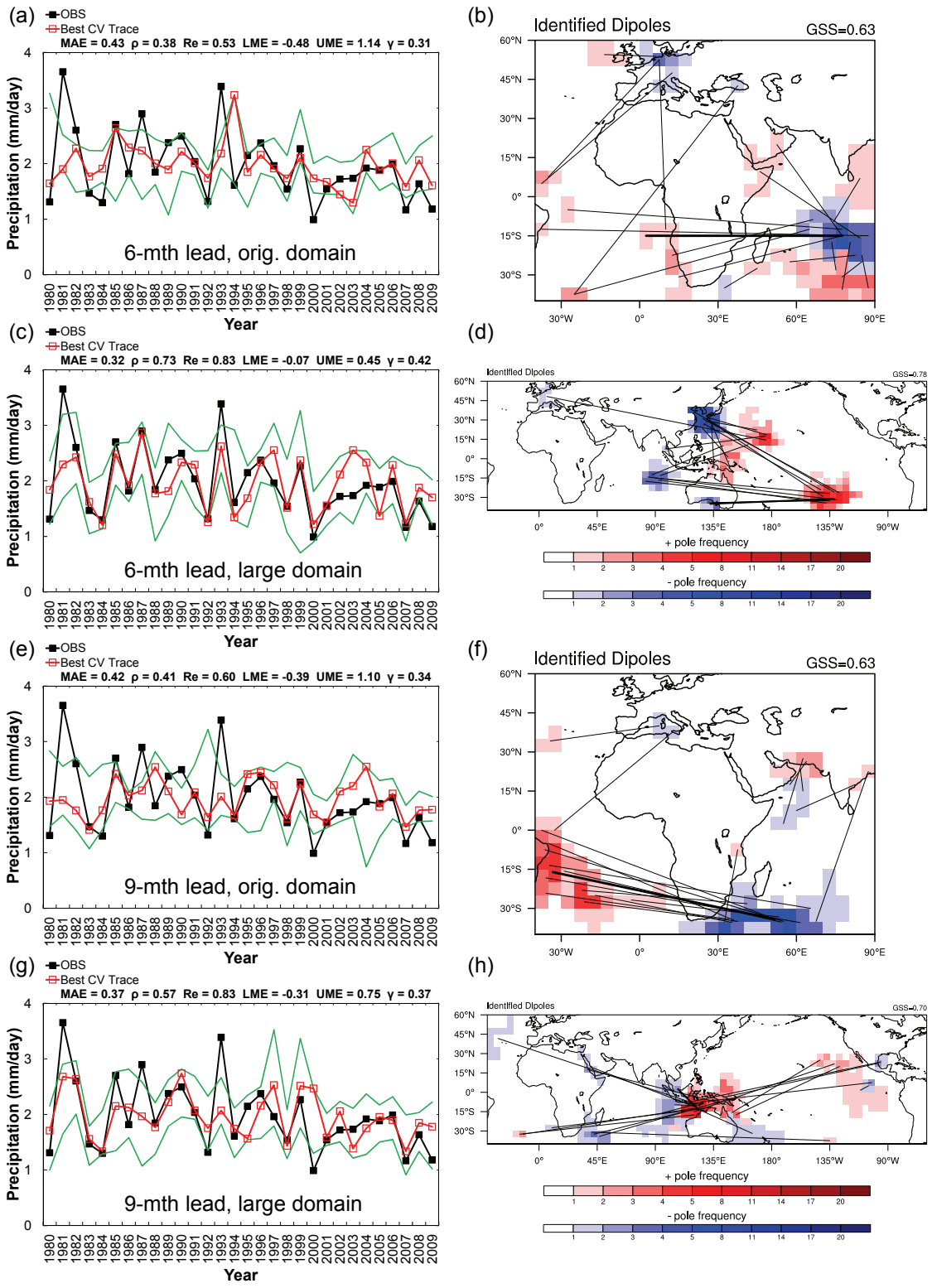



**Figure 5.27:** As in **Figure 5.24**, but for the long rains at Grid 11; results at 6- and 9-month lead times are shown.

*5.3.3 Kiremt Rains at Grid 9*

As expected, the updated forecasting skills for the Kiremt rains at Grid 9 (**Table 5.9**) also exhibit some improvement. The highest Re value increases to 0.700 (from 0.667), and the best CV correlation coefficient increases to 0.665 (from 0.609). Such improvement is attributed to some newly identified dipoles over the Pacific Ocean and their connections with the original dipoles over the Indian Ocean (**Figure 5.28**).

Since the sensitivity analysis carried out in this section clearly indicates that including a larger SSTA domain improves the dipole identification for all three rain types, future execution of the dipole algorithm for other predictands of interests can probably begin with a near-global SSTA domain if no solid mechanism has been understood. Being inclusive of the narrower domains, this larger domain can still identify dipoles that reside within smaller oceanic regions.

**Table 5.9:** As in **Table 5.7**, but for the Kiremt rains at Grid 9.

| Gd_09 (Larger Domain) | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 1 | 0.485 | 0.431 | 0.500 | -1.084 | 1.177 | 0.271 |
| 2 | 0.548 | 0.322 | 0.367 | -1.572 | 1.040 | **0.246** |
| 3 | 0.542 | 0.033 | 0.500 | -1.538 | 1.023 | 0.285 |
| 4 | 0.462 | 0.482 | 0.533 | -1.509 | 1.201 | 0.261 |
| 5 | 0.432 | 0.371 | 0.667 | -1.401 | 0.929 | 0.319 |
| 6 | **0.408** | 0.616 | **0.700** | -1.248 | **0.485** | 0.351 |
| 7 | 0.417 | **0.665** | 0.667 | -1.041 | 0.503 | 0.335 |
| 8 | 0.462 | 0.415 | 0.533 | -1.147 | 0.767 | 0.318 |
| 9 | 0.464 | 0.536 | 0.633 | -1.329 | 0.657 | 0.352 |
| 10 | 0.473 | 0.502 | 0.667 | -0.931 | 0.816 | 0.358 |
| 11 | 0.475 | 0.541 | 0.667 | -0.978 | 0.918 | 0.338 |
| 12 | 0.447 | 0.517 | 0.633 | **-0.848** | 0.795 | 0.353 |

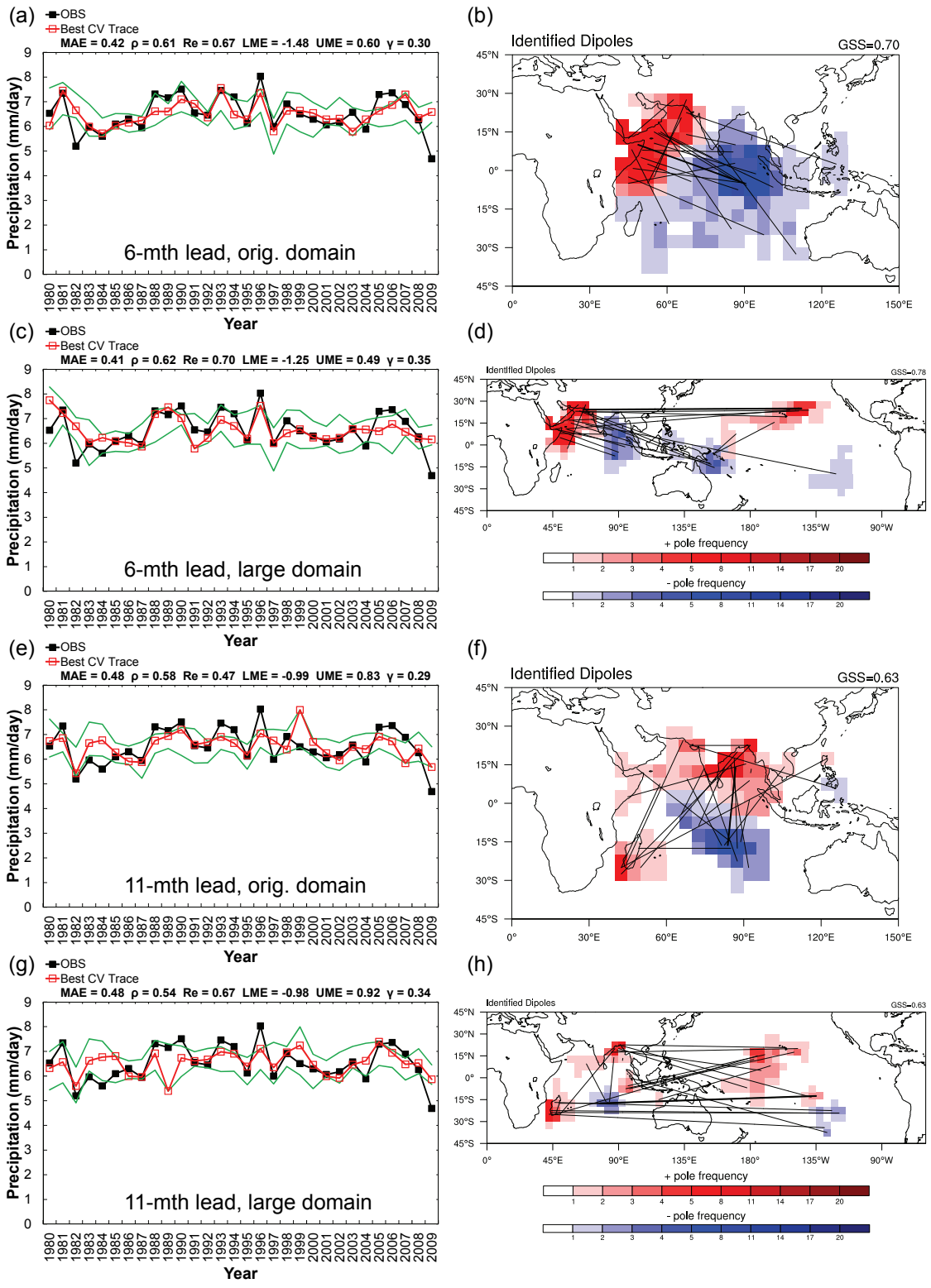



**Figure 5.28:** As in **Figure 5.24**, but for the Kiremt rains at Grid 9; results at 6- and 11-month lead times are shown.

## 5.4 Intercomparison of Forecasting Schemes

As in Chapter 4, a comparison of hindcasting skills generated by different statistical methods is undertaken to verify whether the dipole method is robust. In addition to the dipole algorithm, the following statistical methods are also used. As in the previous chapter, the first alternative method is CCA due to its proven effectiveness and operational use. For a fair comparison, the same SSTA domains used for identifying the dipoles corresponding to all different rain types in East Africa (i.e., the original short- or long-rains domains without expansion as conducted in the last section) are applied to the CCA method. The second method is a simple linear regression scheme, but the predictor is an index of the IOD rather than ENSO since the Indian Ocean is thought to impact more the East African climate. According to Saji et al. (1999) and the information available on the website of Japan Agency for Marine-Earth Science and Technology (http://www.jamstec.go.jp/frsgc/research/d1/iod/), the Dipole Mode Index (DMI), defined as the SSTA gradient between the western equatorial Indian Ocean ($50^{o}$E–$70^{o}$E and $10^{o}$S–$10^{o}$N) and the south eastern equatorial Indian Ocean ($90^{o}$E–$110^{o}$E and $10^{o}$S–$0^{o}$N), is derived from the Kaplan data and adopted as the predictor in the regression method. The last method is climatological (CLM) forecasting. To reiterate, CLM uses a constant forecast based on the long-term average of each rain type and can be an evaluation threshold that justifies if any of the above methods present substantial forecasting skill. Moreover, the schemes of leave-one-out cross-validation and variable lead times consistent with the dipole method are employed to evaluate the best CV hindcasting skills.

After applying these methods to the same case study, the hindcasting skill in terms of $MAE_{CV}$ and $\rho_{CV}$ for all methods are summarized in **Table 5.10**. In general, the dipole method universally outperforms all other statistical methods in both $MAE_{CV}$ and $\rho_{CV}$. While CCA still exhibits weaker but substantial forecasting skills (in comparison with CLM forecasting) for all rain types, IOD regression can only predict the short rains, indicating that the Indian Ocean influences very little the variability of East African rainfall except for the short rains. Further, the patterns of the best lead times among these three methods agree with those identified by the dipole method. For instance, most of the best lead times for the short rains are short or near-concurrent, but those for the T-unimodal rains are quite long. As addressed in Chapter 4, the disadvantages that cause CCA and IOD regression to be inferior to the dipole method are their over-emphasis on either statistical or physical relationships. Specifically, incorporating too many spatial variables risks CCA to become over-fitted and generate artificial skill, while incorporating too few spatial variables prevents IOD regression to flexibly respond to more variable climate in East Africa. However, the dipole method overcomes these disadvantages by achieving a compromise between statistical predictors and is thus more robust characterized by higher forecasting skill for East African rainfall.

**Table 5.10:** Comparisons between the hindcasting skills ( $MAE_{CV}$ , $\rho_{CV}$ ) of the dipole (DP) method and the skills of other operational or common methods.

| Rain Type | Season | Region | Method | | | |
|---|---|---|---|---|---|---|
| | | | DP* | CCA* | IOD* | CLM† |
| Short Rains | ON | Gd_03 | 0.52, 0.71 (1) | 0.63, 0.65 (1) | 0.61, 0.57 (1) | 0.80, *NA* |
| | ND | Gd_06 | 0.83, 0.56 (1) | 0.95, 0.49 (8) | 0.94, 0.53 (2) | 1.10, *NA* |
| | ON | Gd_07 | 0.97, 0.61 (1) | 1.26, 0.32 (1) | 1.14, 0.28 (2) | 1.20, *NA* |
| | ON | Gd_08 | 0.60, 0.62 (1) | 0.77, 0.43 (1) | 0.74, -0.19 (4) | 0.74, *NA* |
| | ON | Gd_10 | 0.90, 0.58 (2) | 1.24, 0.18 (3) | 1.14, 0.01 (3) | 1.20, *NA* |
| | ON | Gd_11 | 0.49, 0.67 (1) | 0.67, 0.54 (1) | 0.65, 0.49 (1) | 0.77, *NA* |
| | ON | Gd_13 | 0.55, 0.59 (7) | 0.72, 0.41 (9) | 0.64, 0.31 (1) | 0.72, *NA* |
| Long Rains | MAM | Gd_03 | 0.44, 0.76 (9) | 0.53, 0.67 (10) | 0.70, -0.30 (6) | 0.69, *NA* |
| | MAM | Gd_06 | 0.62, 0.58 (9) | 0.74, 0.42 (5) | 0.87, -0.25 (1) | 0.85, *NA* |
| | MAM | Gd_07 | 0.57, 0.71 (11) | 0.76, 0.48 (12) | 0.83, 0.08 (9) | 0.85, *NA* |
| | MAM | Gd_08 | 0.47, 0.67 (1) | 0.61, 0.52 (1) | 0.67, -0.03 (10) | 0.67, *NA* |
| | AMJ | Gd_10 | 0.45, 0.60 (12) | 0.60, 0.30 (3) | 0.61, 0.16 (11) | 0.65, *NA* |
| | MAM | Gd_11 | 0.36, 0.51 (2) | 0.46, 0.49 (3) | 0.49, 0.09 (6) | 0.51, *NA* |
| | MAM | Gd_12 | 0.47, 0.49 (5) | 0.63, 0.09 (4) | 0.58, -0.20 (8) | 0.59, *NA* |
| | MAM | Gd_13 | 0.42, 0.55 (3) | 0.60, 0.02 (4) | 0.51, 0.09 (6) | 0.55, *NA* |
| | MAM | Gd_14 | 0.44, 0.51 (4) | 0.56, 0.17 (4) | 0.54, 0.17 (5) | 0.57, *NA* |
| T-unimodal | NDJFMA | Gd_01 | 0.37, 0.49 (10) | 0.47, 0.30 (10) | 0.48, 0.23 (11) | 0.50, *NA* |
| | NDJFMA | Gd_02 | 0.33, 0.72 (12) | 0.47, 0.44 (11) | 0.53, -0.71 (3) | 0.52, *NA* |
| | DJFMA | Gd_05 | 0.49, 0.62 (12) | 0.65, 0.37 (8) | 0.68, -0.29 (12) | 0.68, *NA* |
| Others | AMJJASO | Gd_04 | 0.32, 0.56 (10) | 0.41, 0.21 (5) | 0.42, 0.22 (12) | 0.43, *NA* |
| | JJAS | Gd_09 | 0.42, 0.61 (6) | 0.54, 0.40 (7) | 0.61, 0.00 (12) | 0.61, *NA* |
| | JAS | Gd_12 | 0.33, 0.62 (11) | 0.46, 0.36 (7) | 0.48, -0.26 (5) | 0.48, *NA* |

*: Numbers in parentheses represent the best lead times (in month) corresponding to each season

†: $\rho$ is not available ($NA$) for CLM

## 5.5 Diagnostic Analysis of Large-Scale Variables

As in Chapter 4, diagnostic analysis of additional large-scale variables to SST pertaining to wet or dry composite years can facilitate the understanding of hypothetical mechanisms of specific rain types and clarify any plausible connections with the identified dipoles. According to a specific rainfall series (e.g., the short or long rains), ten wet and ten dry years (out of 30 years) when rainfall is higher and lower than its long-

term 67th and 33rd percentiles are recognized. Because the higher predictability of a certain rain type may reflect more prominent circulation patterns, only representative grids for which the dipole algorithm presents higher forecasting skills are selected to generate composite maps. Following the same selection logic in Section 5.3, these grids include Grids 3 and 11 for both the short and long rains and Grid 9 for the Kiremt rains. Results of diagnostic analysis of the T-unimodal and southern Sudan rains are not discussed in this section because the results of the former rains bear high resemblances to those of the long rains (likely due to some overlapping temporal periods between these rain types), and the hindcasting skills of the latter rains are relatively insignificant.

All composite years associated with different rain types and grids are listed in **Table 5.11**. Because of the same rain types for Grids 3 and 11, high repetitions in composite years are expected. However, three to four different composite years can still be noticed, which may produce certain contrasting patterns that suggest diverse physical linkages. The wet and dry years of the long rains, especially for Grid 3, clearly reiterate the long-term downward trend since most of the wet (dry) years occurred before (after) the 1990s. In contrast, the wet and dry years of the Kiremt rains seem to distribute more evenly over the last three decades.

**Table 5.11:** Wet and dry years (corresponding to different rain types and regions in East Africa) used in composite analysis.

| Region | Season | Composite Years | |
|---|---|---|---|
| | | Wet | Dry |
| Gd_03 | ON | 1982, 1984, 1987, 1989, 1991, 1994, 1997, 2001, 2002, 2006 | 1980, 1981, 1988, 1993, 1996, 1998, 2003, 2005, 2007, 2008 |
| | MAM | 1981, 1982, 1983, 1985, 1986, 1987, 1988, 1989, 1990, 1996 | 1992, 1998, 1999, 2000, 2001, 2002, 2003, 2007, 2008, 2009 |
| Gd_11 | ON | 1982, 1989, 1994, 1997, 2002, 2004, 2006, 2007, 2008, 2009 | 1980, 1983, 1986, 1988, 1991, 1993, 1996, 1998, 2003, 2005 |
| | MAM | 1981, 1982, 1985, 1987, 1989, 1990, 1993, 1995, 1996, 1999 | 1980, 1983, 1984, 1992, 1994, 1998, 2000, 2001, 2007, 2009 |
| Gd_09 | JJAS | 1981, 1988, 1989, 1990, 1993, 1994, 1996, 1998, 2005, 2006 | 1982, 1983, 1984, 1985, 1987, 1995, 1997, 2001, 2004, 2009 |

*5.5.1 Short-Rains Composite Fields (Grids 3 and 11)*

**Figure 5.29** illustrates composite maps for the wet and dry ON years of Grid 3. At the 500mb level, equatorial easterly (westerly) anomalies are pronounced during the wet (dry) years, enhancing (reducing) the zonal moisture transport to East African regions. In northern subtropical areas, an anomalous low (high) centered at the Near East can be found during the wet (dry) years, and this anomalous system is even extended to Eastern Europe during the dry years [**Figure 5.29(b)**]. In essence, the geopotential height anomalies may be conducive to the route of subtropical jet streams, thereby affecting the confluence/diffluence of equatorial flows over Maritime Continent. In the Southern Hemisphere, while a moderate anomalous high is situated over Australian regions during the wet years, an extraordinary tripole structure of geopotential height anomalies can be observed during the dry years. These geopotential height anomalies may dominate the

condition of the South Indian Ocean tradewinds and alter the equatorial flows. At the surface level, a clearer zonal pressure gradient that induces easterly or westerly anomalies can be seen. Overall, the Walker-like circulation found, which comprises the weakened Wyrki Jet (i.e., anomalous equatorial easterlies in the upper ocean), anomalous ascending motion over East Africa, anomalous equatorial westerlies at the upper-troposphere (200mb, not shown), and descending motion over Indonesia, is consistent with analyses in many existing studies (e.g., Hastenrath, 2007; Hastenrath et al., 2007, 2011).

**Figure 5.30** shows another group of composite maps for the wet and dry ON years of Grid 11 and reveals similar characteristics with several modifications. For instance, at the 500mb level, high anomalies in the North Asia are more enhanced and intrude farther in the Black Sea area, separating the subtropical low anomalies in the Near East during the wet years. Besides, the tripole structure in the Southern Hemisphere diminishes and transforms into enhanced low anomalies that emanate from the south coast of Australia and elongate to the south coast of Africa during the dry years. Similarly, at the surface level, some more enhanced pressure anomalies over extratropical areas (e.g., the contrasting patterns between pressure anomalies at Eastern Europe in the wet years and those in the dry years) and some modified patterns over tropical regions (e.g., the diminished low anomalies in the eastern Indian Ocean during the dry years) can be found. Despite all these enhancements or modifications, the equatorial easterly/westerly anomalies, proven to be of great importance for the East African short rains, still remain in the composite maps for Grid 11. Even though the main driving force is nearly identical, some little distinction of the anomalous easterlies at the downstream side

during the wet years can be recognized: a slightly stronger northerly element near the location of Grid 11 and the coastal area trailed by enhanced coastal low anomalies. This distinction of the wind curvature may help to explain why the major positive poles for Grid 11 are more emphasized at the Arabian Sea and the coast of Somalia [see **Figure 5.10(a)**]. Moreover, the different subtropical (and extratropical) forcing as mentioned can also partially contribute to the changes in rainfall amounts.

To further examine whether the identified dipoles are legitimate as well as to elucidate the evolution of the oceanic field, wet-minus-dry SSTA composite maps for Grid 3 at different lead times are generated in **Figure 5.31**. The figure shows that the short rains at Grid 3 exhibit a clear relationship with the ENSO transition from La Niña patterns at long lead times (8–12 months) to El Niño patterns at short lead times (1–6 months). Such relationship has been found in many other studies (e.g., Janowiak, 1988; Ogallo, 1988; Mutai et al., 1998), yet some other researchers argue that although ENSO is pronounced when East Africa experiences wet or dry events, it plays no more than an indirect role in modulating the Indian Ocean (more specifically, the Indian Ocean Dipole), regarded as the more predominant forcing of the short rains (e.g., Goddard and Graham, 1999; Latif et al., 1999; Black et al., 2003).

A different statement associated with ENSO can possibly be made according to the dipole results in this study. At the 1-month lead time, the proposed algorithm identified several oceanic areas as the components of dipoles, including the southeastern to central Indian Ocean and the complementary ENSO areas (e.g., negative poles in the western and

southern Pacific Ocean), other than the "tongue" area in the eastern tropical Pacific [see **Figure 5.24(d)**]. On the other hand, at the 9-month lead time, although the dipoles contain some cold-tongue areas, the main active SSTA regions are the southern Pacific and a large portion of the Indian Ocean [see **Figure 5.24(h)**]. In general, the algorithm can identify dipoles at consistent locations with the patterns of the composite maps, but it does not emphasize the signature ENSO pattern (i.e., the tongue area) for the following reasons: 1) Some historical short-rains events that incorrectly respond to ENSO may drastically worsen forecasting skill (e.g., wet events in 1984, 1989, and 2001 would wrongly be predicted as dry events because of La Niña signals). 2) Consequently, dipoles situated in the eastern tropical Pacific are ruled out from the final dipole list, in which only a limit number $(\xi = 20)$ of more skilled dipoles situated in other active oceanic areas are retained. Therefore, these results state that rather than being the direct or indirect forcing of the short rains, ENSO (at least the principal tongue area) basically exhibits lesser "forecasting utility" for the short rains. Nevertheless, since some Pacific areas are still identified as the components of dipoles, the contributions of the Pacific Ocean to the forecasting of the short rains cannot be neglected.

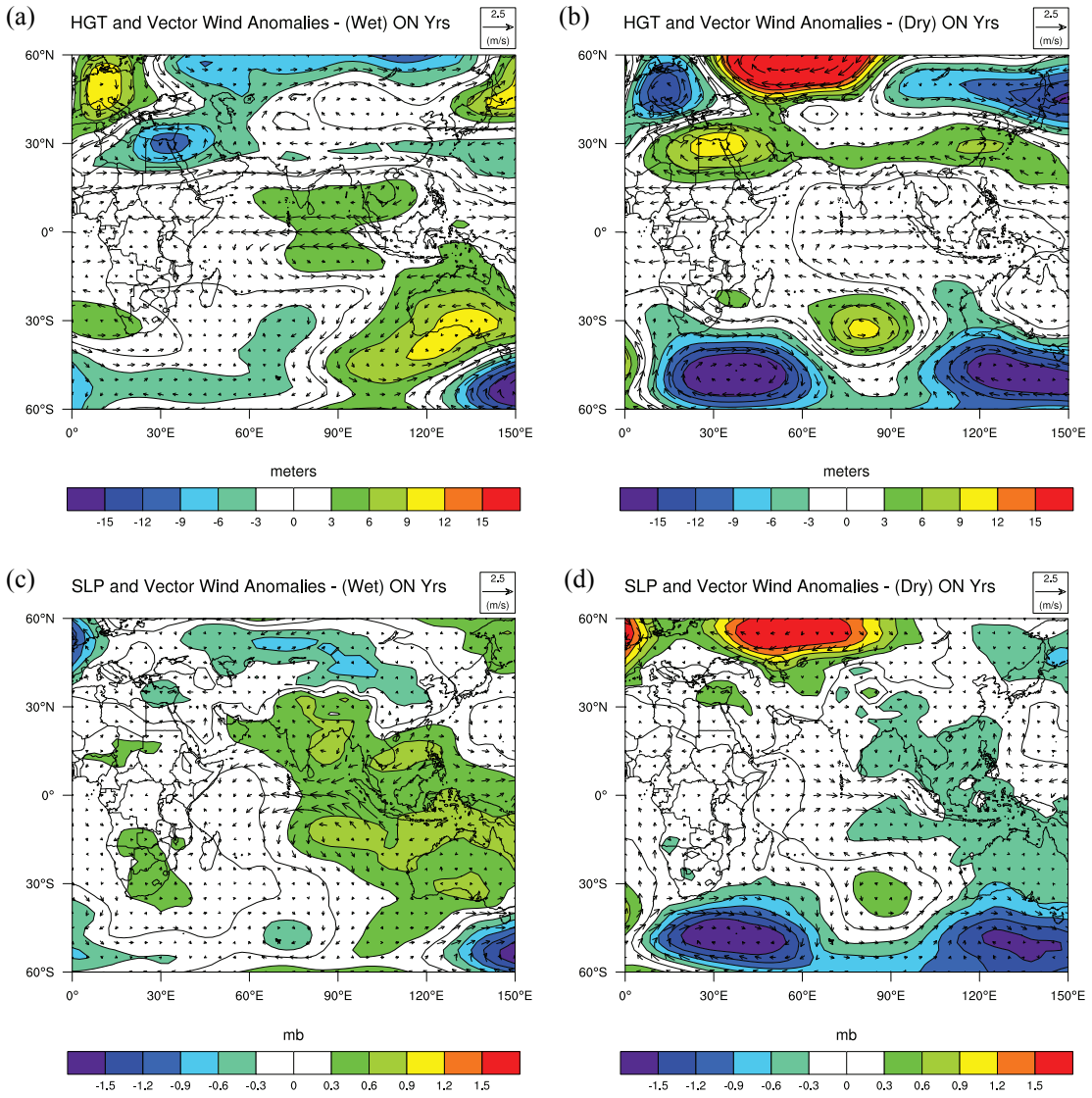



**Figure 5.29:** ON composite maps of NCEP/NCAR Reanalysis data for wet [(a) and (c)] and dry [(b) and (d)] years of Grid 3. The upper panel is the 500mb geopotential height and vector wind anomalies and the lower panel is the SLP and vector wind anomalies.

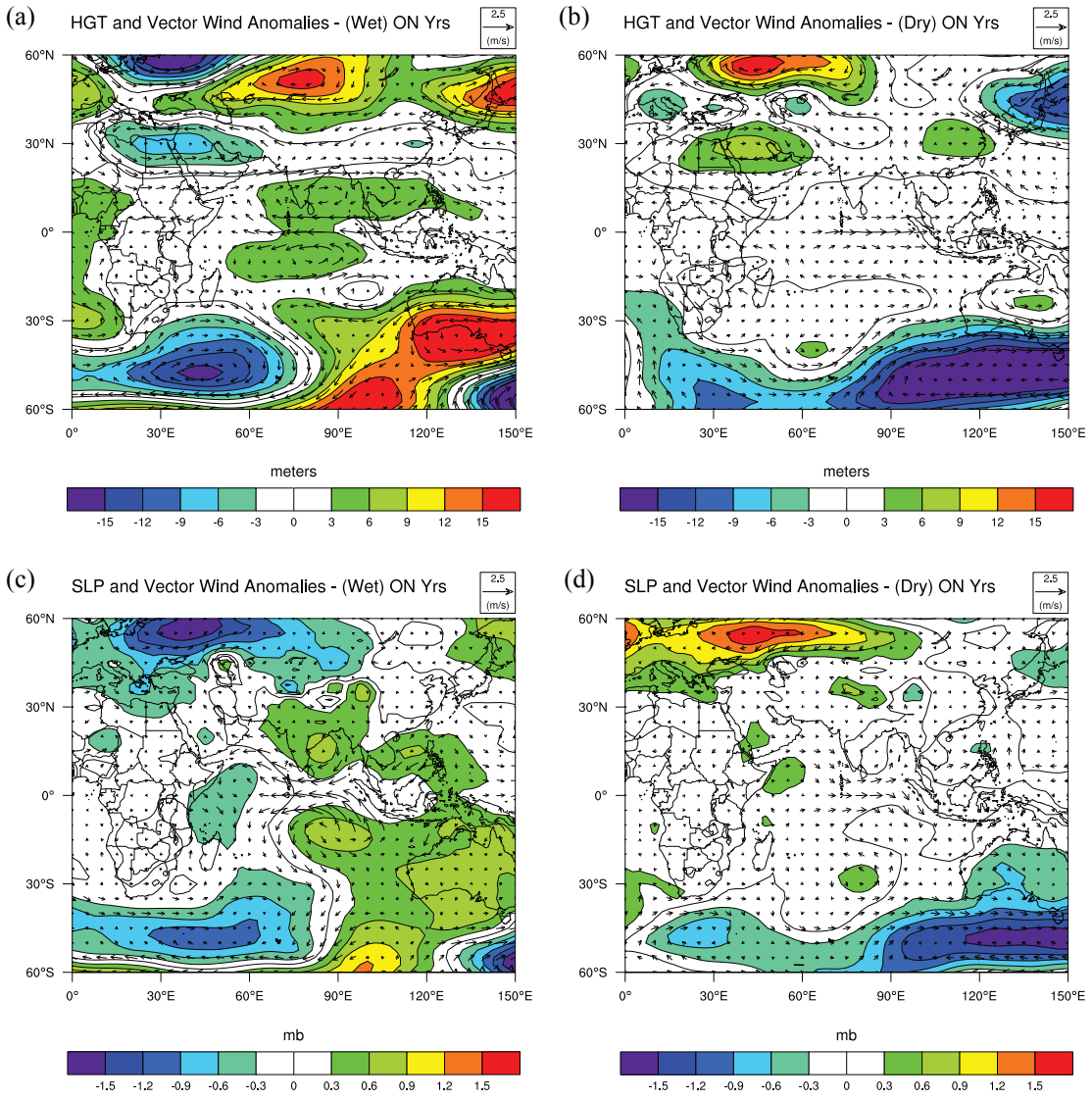



**Figure 5.30:** As in **Figure 5.29**, but based on composite years of Grid 11.

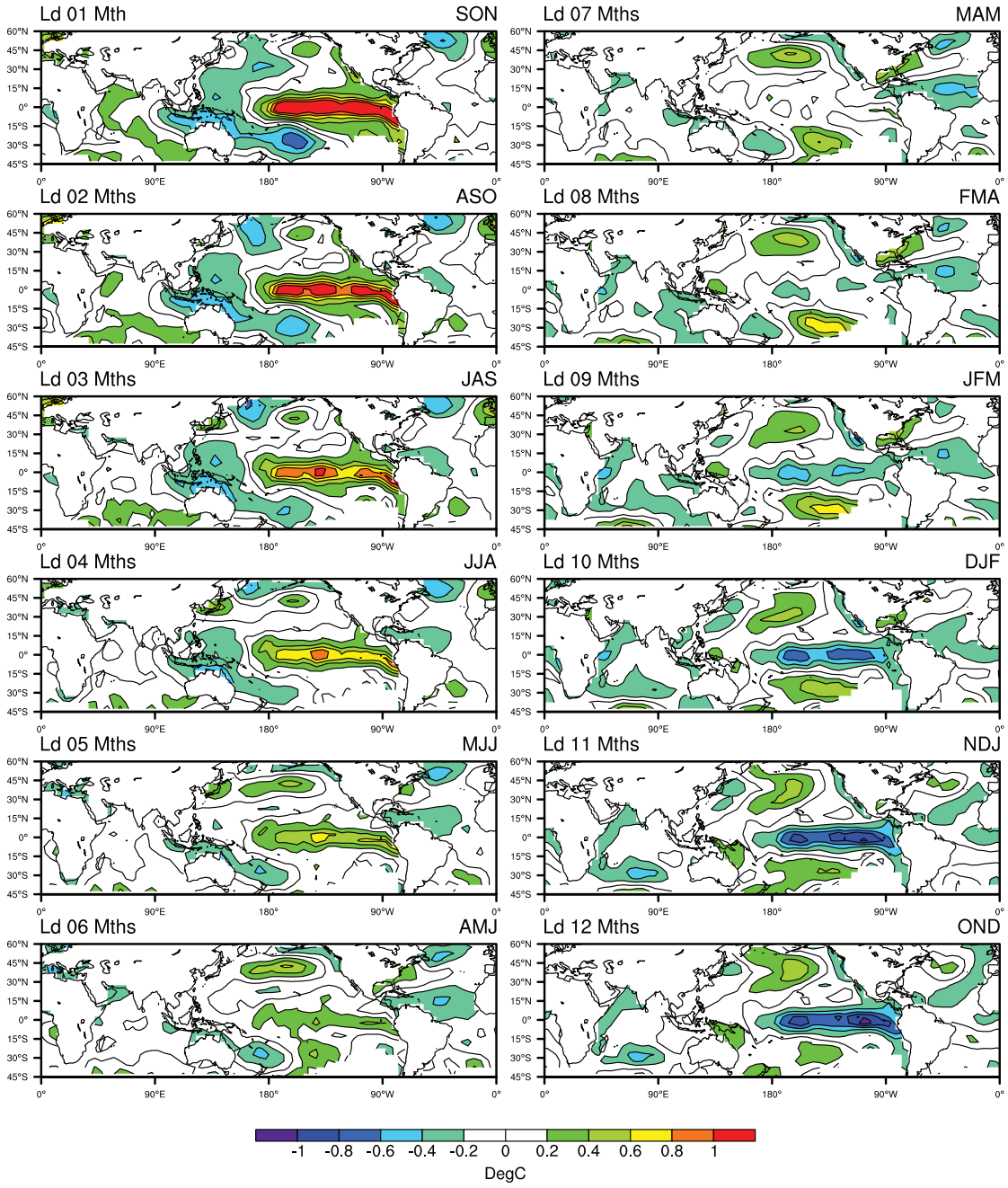



**Figure 5.31:** Evolution of SSTA composites in wet-minus-dry (ON) years of Grid 3.

*5.5.2 Long-Rains Composite Fields (Grids 3 and 11)*

With the same aim of understanding the underlying climatic forcing, composite maps in accordance with the wet and dry long-rains years of Grids 3 and 11 are plotted in **Figures 5.32** and **5.33**, respectively. Both figures reveal almost identical patterns near the tropical regions. For example, at the 500mb level, cyclonic anomalies in West Africa and associated notable westerly anomalies that carry more unstable, moist air from the Congo Basin and the Gulf of Guinea can be found during the wet years. In contrast, reversed patterns that impede the incursion of moist air into East Africa are shown during the dry years. However, at the surface level, neither significant pressure nor wind anomalies can be found over the tropical regions.

More pronounced and contrasting patterns seem to emerge over the extratropical regions. For instance, during wet years, the aforementioned 500mb cyclonic anomalies in West Africa actually originate from more enhanced low anomalies from the North Atlantic Ocean and may traverse across North Africa and all the way to South Asia [**Figure 5.33(a)**]. Meanwhile, high anomalies can be observed in the Eastern and Northern Europe, though the anomalies are milder for Grid 11. During the dry years, partially reversed patterns can be seen with some differences between the patterns for Grid 3 and those for Grid 11. While positive geopotential height anomalies extensively occupy the North Atlantic Ocean and Europe for Grid 3, such anomalies are restricted within the oceanic area for Grid 11; and the pattern over Europe (and parts of the Mediterranean Sea and the Near East) is replaced with low anomalies. At the surface level, extensive high (low) anomalies that emanate from northern Russia and Siberia during the wet (dry) years

can also be observed except that the low anomalies migrate westward for Grid 11. In addition, some far-south patterns in the South Atlantic and Indian Oceans are shown in the composite maps.

Nevertheless, unlike the well-established relationship between the long rains and the westerly anomalies from the Congo Basin, many of these extratropical patterns are new or only partially consistent with results shown in previous studies (e.g., Camberlin and Philippon, 2002; Camberlin and Okoola, 2003) that conducted similar composite analysis but using data in earlier decades (from the 1960s to 1990s). Since some other researchers suggested that the downward trend of the long rains in recent decades is associated with the warming western tropical Pacific (Williams and Funk, 2010) and even exacerbated by similar forcing after the late 1990s (Lyon and DeWitt, 2012), it can be hypothesized that the inconsistent patterns may be attributed to a general change in large-scale circulation.

To examine if the SSTA field also adheres to the inconsistency and to illustrate its connections with the identified dipoles at different lead times, the evolution of wet-minus-dry SSTA composite maps for Grid 3 is shown in **Figure 5.34**. Firstly, in contrast to the same figure for the short rains, ENSO is not eminent at short lead times and only display weak La Niña signals around the Nino 1+2 region (the extreme eastern tropical Pacific) at lead times longer than seven months for the long rains. This weak and negative correlation (i.e., negative SSTA in the eastern Pacific may lead to positive long-rains anomalies, and vice versa) is consistent with previous studies (e.g., Ogallo, 1988; Indeje et al., 2000). Secondly, the correspondence between the components of dipoles

and the prominent oceanic areas in the composite maps is reasonable. For instance, at the 9-month lead time, the proposed algorithm successfully identified the La Niña signals accompanied with several additional areas, such as the cold Mediterranean and Black Seas and the warm South Atlantic Ocean [see **Figure 5.26(d)**]. At the 5-month lead time, in line with the weaker La Niña signals, the algorithm excluded patterns in the eastern Pacific and identified dipoles consisting of relatively more active oceanic areas, including the cold northern subtropical Pacific, the cold Red Sea, and the cold North and warm South Atlantic Oceans [see **Figure 5.26(h)**].

Lastly, even though the identified dipoles seem to be valid and may further link to the atmospheric patterns as mentioned (because the SSTA composites fairly persist from long to near-concurrent lead times), the dipoles as well as the SSTA composites at the far side of the Pacific are likely to contradict the findings from Camberlin and Philippon (2002) and Camberlin and Okoola (2003). Their work indicated that the warm South Atlantic Ocean and the cold Mediterranean Sea may delay the onset of the long rains and then result in reduced total rainfall amounts, which is generally opposite to the findings of this study based on data in more recent decades. Thus, the SSTA composite analysis supports the earlier statement that the change in large-scale circulation is possibly universal, causing a climate regime shift and altering the way extratropical patterns influence the long rains.

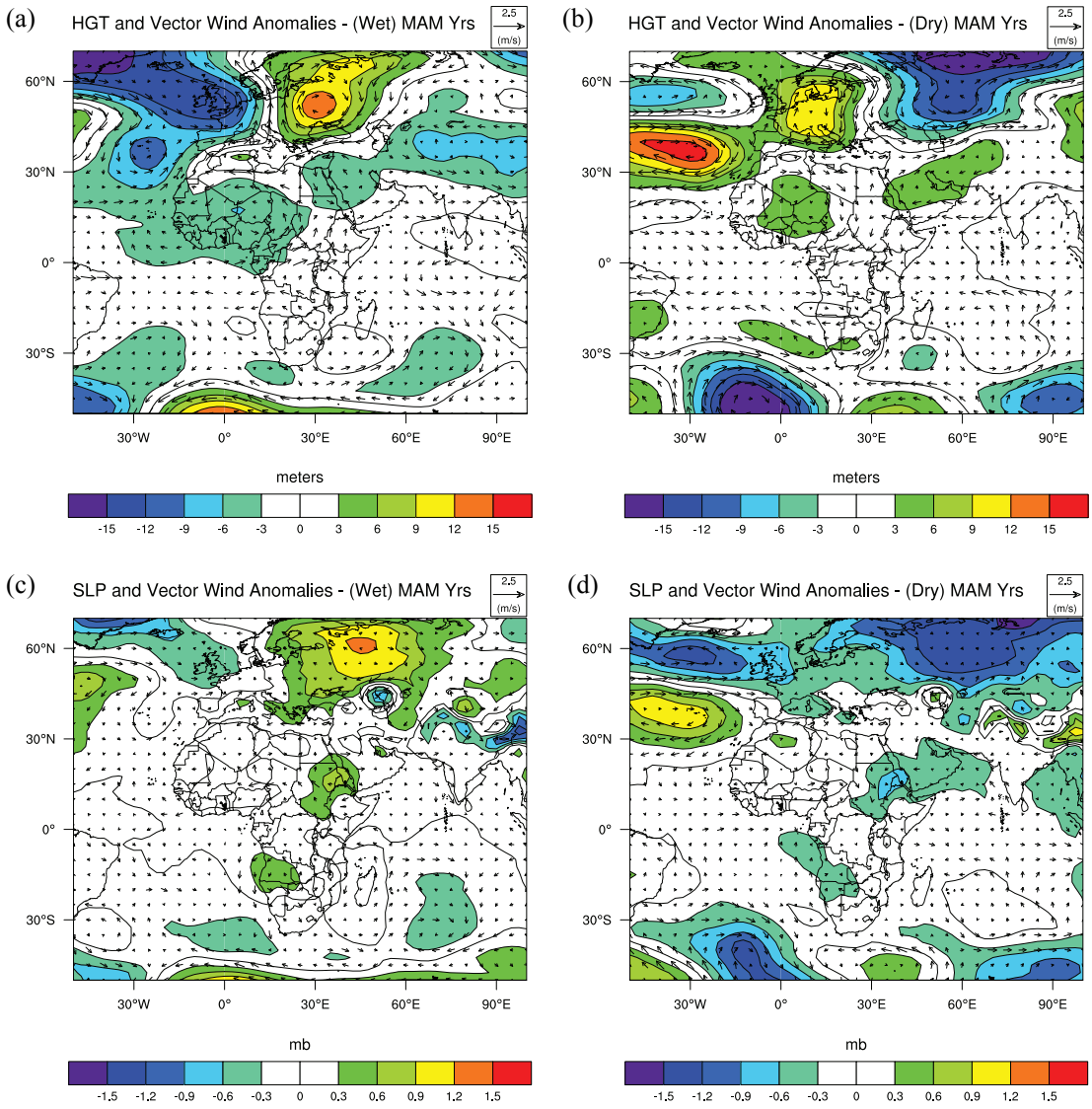



**Figure 5.32:** As in **Figure 5.29**, but for MAM composite maps of Grid 3.

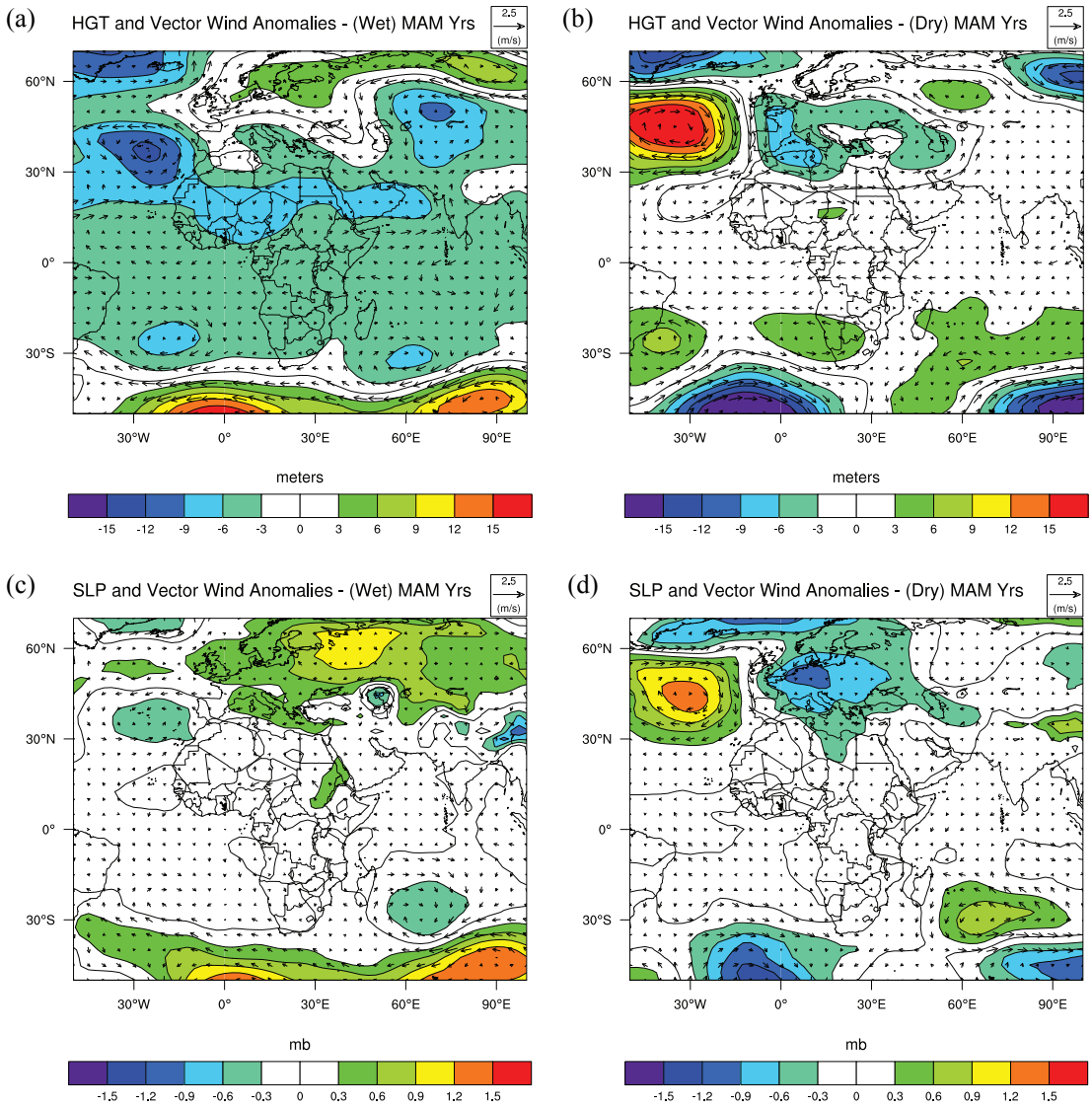



**Figure 5.33:** As in **Figure 5.32**, but based on composite years of Grid 11.

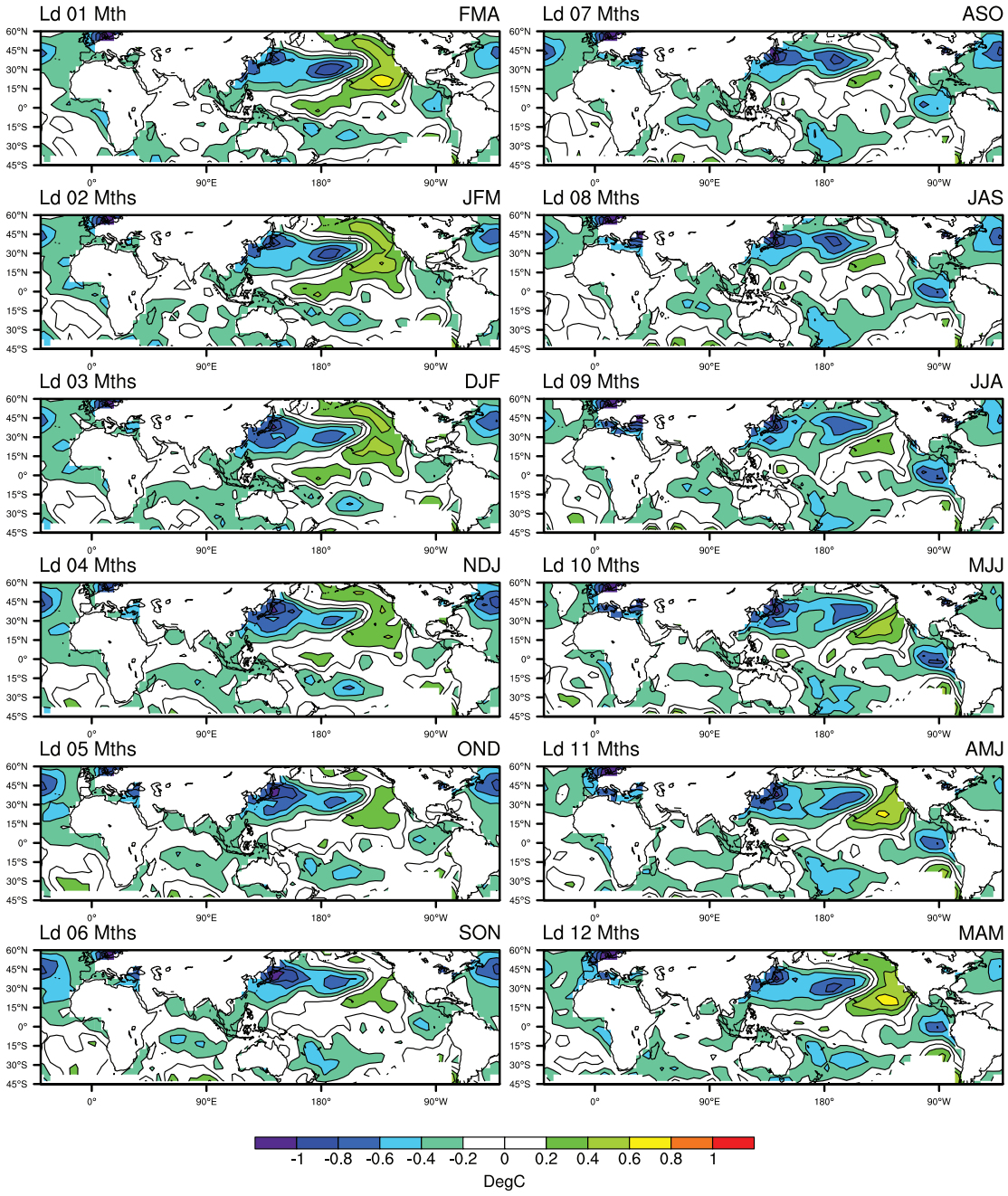



**Figure 5.34:** Evolution of SSTA composites in wet-minus-dry (MAM) years of Grid 3.

*5.5.3 Kiremt-Rains Composite Fields (Grid 9)*

The last group of composite maps at mid-troposphere and surface levels in accordance with the wet and dry Kiremt-rains years for Grid 9 is plotted in **Figure 5.35**. During the dry years, 500mb easterly and northeasterly anomalies that appear at the coast of East Africa and the Arabian Sea indicate the weakened Somali Jet and Southwest Monsoon, thereby discouraging moisture supply from the Indian Ocean. Meanwhile, westerly anomalies can also be found at the western tropical Pacific, implying the atypical Walker circulation over the Pacific Ocean and possible El Niño signals. However, during the wet years, these wind anomalies only exhibit a weak reversal. This suggests that some other mechanisms at other atmospheric levels, such as the Tropical Easterly Jet (TEJ; Segele et al., 2009; Diro et al., 2011a) at high-troposphere, may account for complementary variability of the Kiremt rains. Indeed, the association of the TEJ with the Kiremt rains, such that the enhanced (weakened) TEJ at 200mb may link to wet (dry) spells, is also supported by this analysis (not shown). Nevertheless, other low-level wind anomalies found in previous studies [e.g., East Africa low-level jet or African easterly jet described by Diro et al. (2011a)] are not found and do not appear to be significant in this analysis. Because most of the previous studies only used data prior to 2000, such inconsistency may be attributed to the same suggestion obtained from the last section that the large-scale circulation has experienced some shifts since the late 1990s.

Beyond the tropical regions, wave-train-like geopotential height anomalies appear at the Northern Hemisphere from Europe to Northeast Asia during wet years, yet elongated negative geopotential height anomalies traverse the same vast area during the dry years.

More notably, in the South Indian Ocean, contrasting patterns of low and high anomalies during the wet and dry years can be found at both the 500mb and surface levels. This may be associated with the Southern Hemisphere Annular Mode or the "Antarctic Oscillation" (Gong and Wang, 1999; Thompson and Wallace, 2000), which shows dramatic impacts on high- to mid-latitude climate through the modulation of zonal jet streams. However, how this newly discovered annular mode influences the Kiremt rains is unclear and requires further investigation.

In addition to the migration of the ITCZ, the wind anomalies, and the possible extratropical forcing, the teleconnection impacts of SSTA (e.g., ENSO) on the Kiremt rains have also been corroborated by numerous studies with some forecasting implications (Gissila et al., 2004; Segele and Lamb, 2005; Block and Rajagopalan, 2007; Korecha and Barnston, 2007; Segele et al., 2009; Diro et al., 2011a). To examine if ENSO signals are still substantial after including the latest data and to further depict how sustainable such impacts can be traced backward at lead times, the evolution of wet-minus-dry SSTA composite maps pertaining to Grid 9 are generated in **Figure 5.36**. Consistent with previous studies and the wind anomalies at the western tropical Pacific previously outlined, the figure reveals a La Niña pattern at the 1-month lead time. This indicates the cold (warm) eastern tropical Pacific may result in wet (dry) spells in the Kiremt season at Ethiopia. However, as the lead time increases to four months or longer, the La Niña signal becomes increasingly insignificant.

The La Niña signal was indeed identified by the dipole algorithm as one of the predictor components at the 1-month lead time (not shown). Nevertheless, the forecasting skill produced by the ENSO-related dipoles seems to be weaker than that produced by other significant dipoles at longer lead times (see **Table 5.9**). For example, as shown in the composite map, warming signals that emerge from the Arabian Sea, the coast of Somalia, and the central Pacific Ocean along with a cold signal at the Coral Sea were identified as the more skillful dipole predictors at the 6-month lead time [see **Figure 5.28(d)**].

Moreover, both the composite map and the identified dipoles at very long lead times (e.g., eleven and twelve months) express Central-Pacific El Niño-like signals, which may be worthy of further investigation. All of these results imply that the physical connection between the variation at the principal ENSO region and that of different rain types in East Africa is very subtle. Establishing a predictor-predictand relationship relying on composite analysis may often lead to an ENSO-driven model (since it statistically accounts for greater SST variance), which is sometimes risky or even faulty because other informative oceanic precursors at different lead times may be ignored.

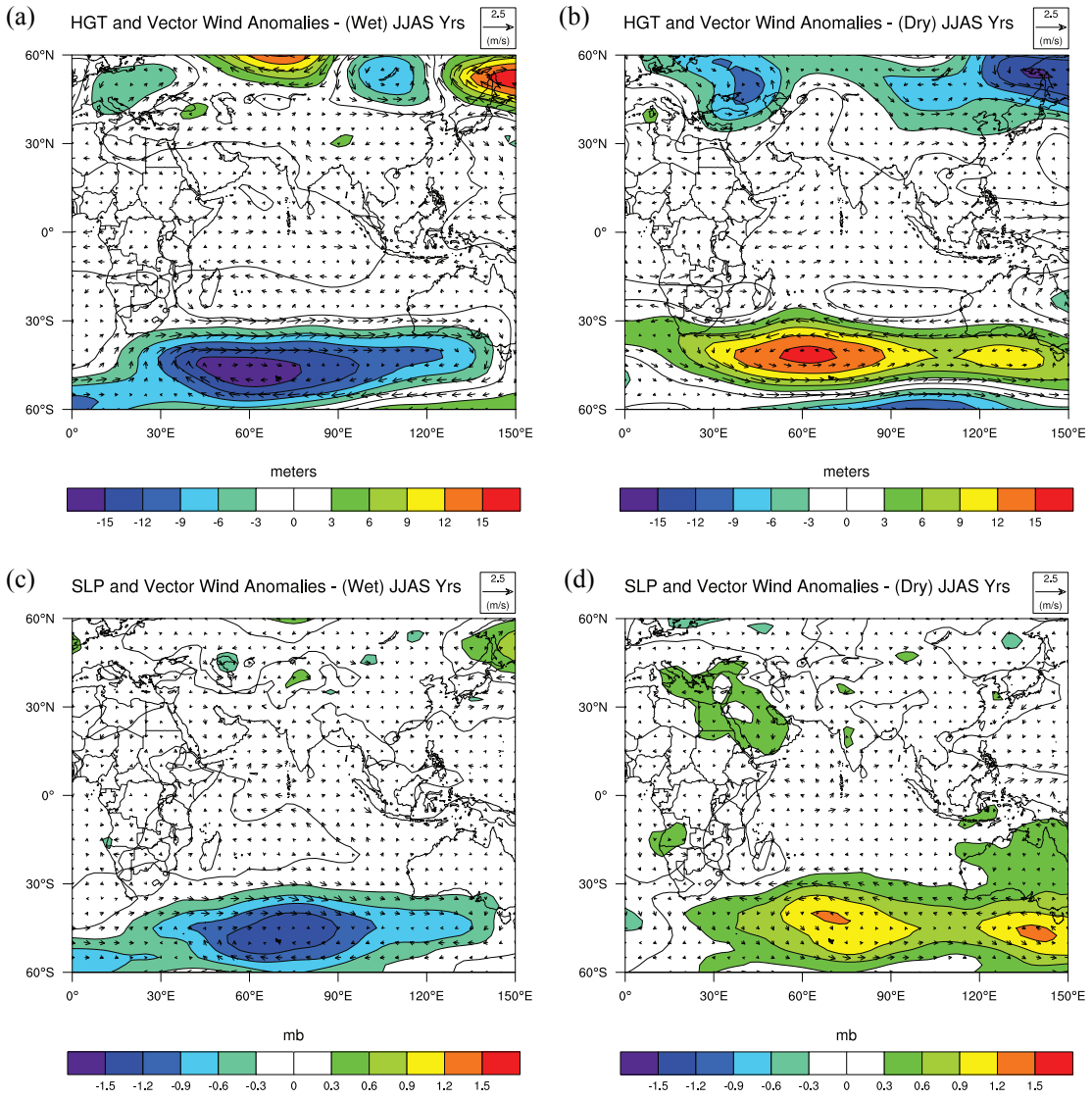



**Figure 5.35:** As in **Figure 5.29**, but for JJAS composite maps of Grid 9.

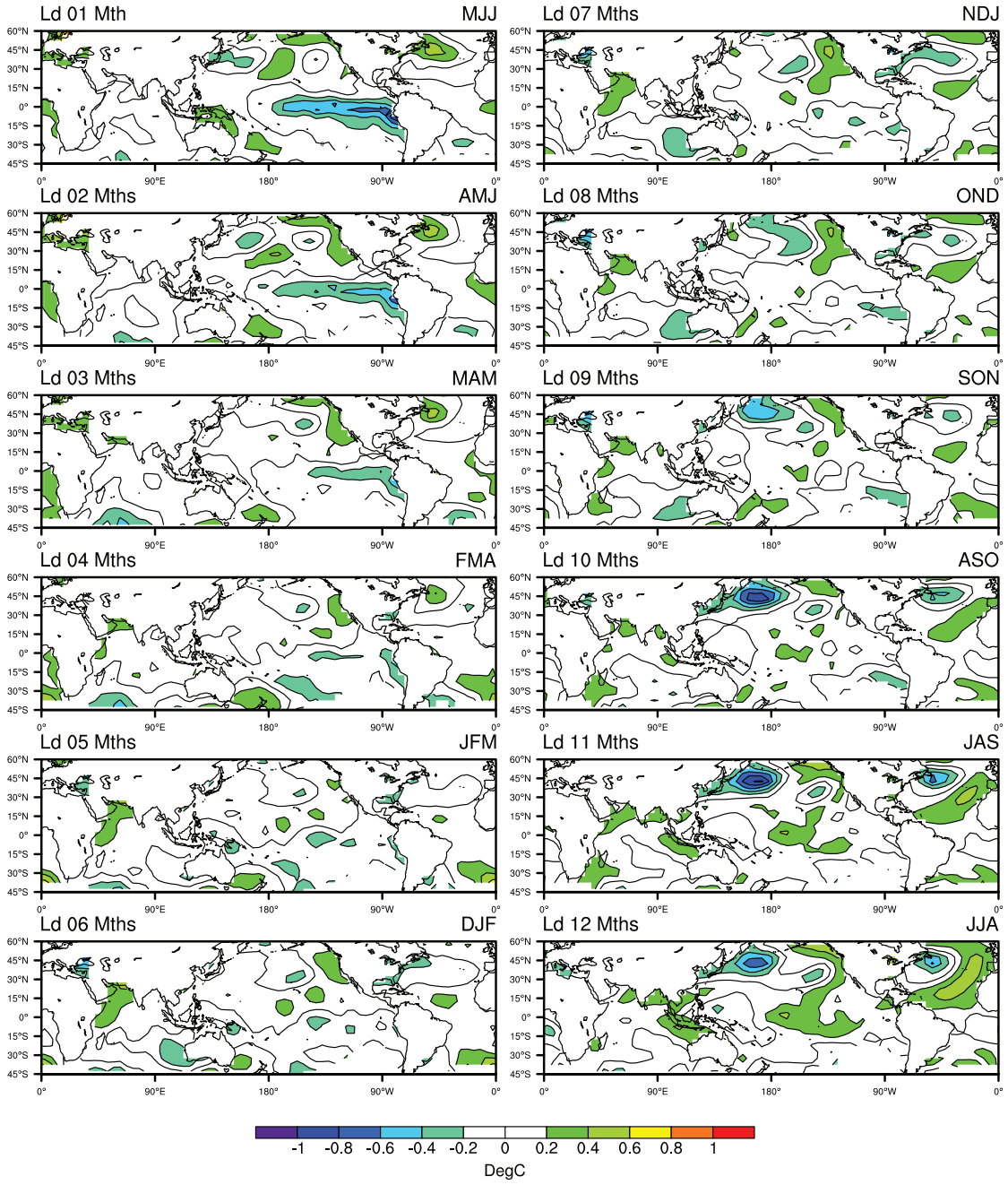



**Figure 5.36:** Evolution of SSTA composites in wet-minus-dry (MAM) years of Grid 9.

## 5.6 Demonstration of 2012 Forecasts

An operational forecasting framework for several types of seasonal precipitation in East Africa is carried out in this section. Similar to the framework introduced in Chapter 4, the best dipole predictors at specific lead times are selected according to the highest Re value for each rain type. Afterwards, these predictors are used to prepare the 2012 forecasts with the newest available SSTA data (until January 2012). Again, it should be noted that the SSTA data availability may impede the selection of dipole predictors from the best lead times based on the hindcasting experiment. This may considerably affect the short-rains forecasts because the strongest correlations with SSTA dipoles are found at short lead times. Thus, for some rain types with relatively low Re values in the following analysis should be updated when new SSTA data become available.

**Table 5.12** summarizes the forecasting results of 2012 seasonal precipitation in East Africa and the associated statistics of the long-term climatology (derived from 1980 to 2009). To reveal the characteristics of forecasting distribution, histograms and boxplots of the forecasting results are also plotted in **Figures 5.37** to **5.39** and **Figures 5.40** to **5.41**, respectively. In addition, the forecasting results for Grids 3, 9, and 11 are generated by dipole predictors identified from the enlarged SSTA domain as described in Section 5.3.

The 2012 short-rains forecasts indicate that an average amount of precipitation is expected for most of the regions except for Grids 10 and 11, where the forecasts appear to have higher mean values. The associated forecasting distribution for either Grids 10 or 11 is more concentrated on higher precipitation rates. Nevertheless, all of the short-rains

forecasts, mainly based on the identified SSTA dipoles in NDJ or OND (i.e., with 11- or 12-month lead times), cannot provide reliable enough information (the highest Re value is only about 63%). According to the previous hindcasting experiment, more reliable forecasts of the short rains should require updated SSTA data (based on either observations or SSTA forecasts) in the near short-rains season.

Unlike the short-rains forecasts, because more pertinent SSTA data is available, the 2012 long-rains forecasts generally have higher reliability (Re values vary from 53–87%). Grounded on the high reliability, the forecasts are more definite and suggest that East Africa will most likely experience another dry season. This also implies that the long-term downward trend of the long rains may persist. A high percentage of ensemble traces presents forecasts lower than the 33rd percentile of climatology of each grid, especially for Grids 3, 8, and 12, where the percentage is as high as 55–95%. The minimums of the forecasts for these grids are even lower than the 30-year records. However, the dry conditions are milder for several coastal grids (e.g., Grids 6, 10, 11, and 13) located at the eastern side of the Great Rift Valley, suggesting that the dry forecasts for the highlands may result from a lack of moist westerlies from the Atlantic Ocean.

Regarding the active Tanzania unimodal rains, the forecasts show some mixed conditions with moderate to high reliability (Re values range from 67–77%). While the near-normal or possibly higher-than-normal forecast is presented at the coastal region (i.e., Grid 5), near-normal to below-normal forecasts are shown for the western parts of Tanzania (i.e., Grids 1 and 2). The driest forecast appears at Grid 2, where 40% of the ensemble traces

are lower than the 33rd percentile of climatology. The relatively dry and wet forecasts are consistent with those for the long-rains-affected grids in terms of geographic correlations: more precipitation in coastal regions (Grids 5 and 6) and less precipitation in highland regions (Grids 2 and 3).

The 2012 forecasts for the southern Sudan and Ethiopian Kiremt rains indicate normal and below-normal conditions, respectively. Whereas the forecasting distribution of the southern Sudan rains is mostly concentrated on the long-term mean, that of the Kiremt rains is more concentrated on lower precipitation rates. Moreover, 25–45% of the ensemble traces for the Kiremt rains exhibit forecasts lower than the 33rd percentile of climatology.

Overall, the year-round outlook for seasonal precipitation in East Africa is relatively dry, which also agrees with the official outlook issued by the Famine Early Warning Systems Network (see www.fews.net). Even though an above-average short-rains season in 2011 has partially relieved stressed and famine-stricken areas, the follow-up shortage in the long rains is ominous, and drought preparedness and amelioration strategies should continue.

**Table 5.12:** Summary of 2012 seasonal precipitation forecasts in East Africa.

| Rain Type | Statistics | Grid Number | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 3* | 6 | 7 | 8 | 10 | 11* | 13 | | |
| Short Rains | Max | 4.62 | 4.33 | 4.41 | 3.22 | 3.74 | 3.10 | 1.99 | | |
| | Mean | 3.58 | 3.39 | 3.55 | 2.52 | 2.28 | 2.12 | 1.11 | | |
| | Min | 2.48 | 2.48 | 2.53 | 2.10 | 1.29 | 1.08 | 0.52 | | |
| | Re | 0.63 | 0.50 | 0.60 | 0.53 | 0.57 | 0.63 | 0.47 | | |
| | LT Max | 6.16 | 6.58 | 9.60 | 6.60 | 8.14 | 5.75 | 3.41 | | |
| | LT Mean | 3.77 | 3.25 | 3.75 | 2.32 | 2.15 | 1.69 | 1.20 | | |
| | LT Min | 2.00 | 1.16 | 1.91 | 1.20 | 0.49 | 0.46 | 0.19 | | |
| **Rain Type** | **Statistics** | **Grid Number** | | | | | | | | |
| | | 3* | 6 | 7 | 8 | 10 | 11* | 12 | 13 | 14 |
| Long Rains | Max | 5.77 | 4.53 | 4.13 | 3.22 | 3.17 | 2.82 | 2.49 | 1.77 | 1.42 |
| | Mean | 4.06 | 3.92 | 3.06 | 2.62 | 2.21 | 1.87 | 1.64 | 1.21 | 0.76 |
| | Min | 2.64 | 3.40 | 1.84 | 1.49 | 1.24 | 1.01 | 0.71 | 0.77 | 0.23 |
| | Re | 0.87 | 0.60 | 0.77 | 0.73 | 0.70 | 0.83 | 0.53 | 0.70 | 0.63 |
| | LT Max | 6.84 | 5.70 | 6.55 | 5.31 | 3.93 | 3.65 | 3.50 | 3.72 | 2.87 |
| | LT Mean | 4.80 | 4.15 | 3.60 | 3.31 | 2.32 | 1.96 | 1.84 | 1.32 | 1.14 |
| | LT Min | 3.04 | 2.11 | 1.30 | 1.64 | 1.00 | 0.99 | 0.78 | 0.11 | 0.07 |
| **Rain Type** | **Statistics** | **Grid Number** | | | | | | | | |
| | | 1 | 2 | 5 | 4 | 9* | 12 | | | |
| T-Unimodal | Max | 5.68 | 4.53 | 6.11 | 3.99 | 6.74 | 2.58 | | | |
| (Gds 1, 2, 5) | Mean | 5.04 | 4.13 | 5.37 | 3.35 | 6.25 | 2.20 | | | |
| S-Sudan | Min | 4.63 | 3.41 | 4.52 | 2.90 | 5.65 | 1.57 | | | |
| (Gd 4) | Re | 0.67 | 0.77 | 0.70 | 0.67 | 0.67 | 0.67 | | | |
| Kiremt | LT Max | 6.30 | 5.79 | 7.11 | 4.50 | 8.03 | 4.04 | | | |
| (Gds 9, 12) | LT Mean | 5.17 | 4.35 | 5.26 | 3.39 | 6.52 | 2.37 | | | |
| | LT Min | 4.07 | 3.07 | 4.01 | 2.19 | 4.69 | 1.30 | | | |

*: dipoles used to generate associated forecasts are identified from a larger SSTA domain

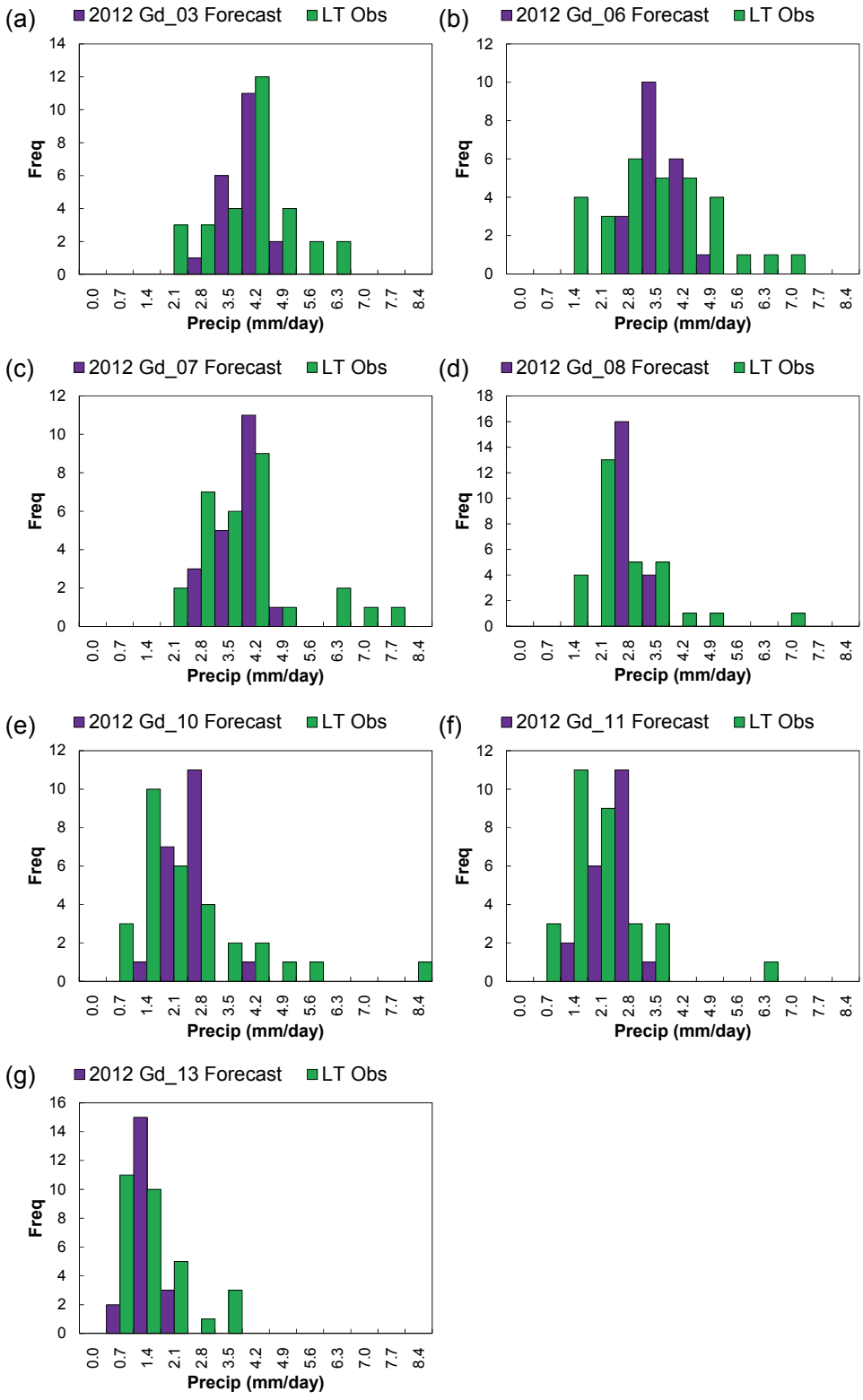



**Figure 5.37:** Histograms of 2012 short-rains forecasts.

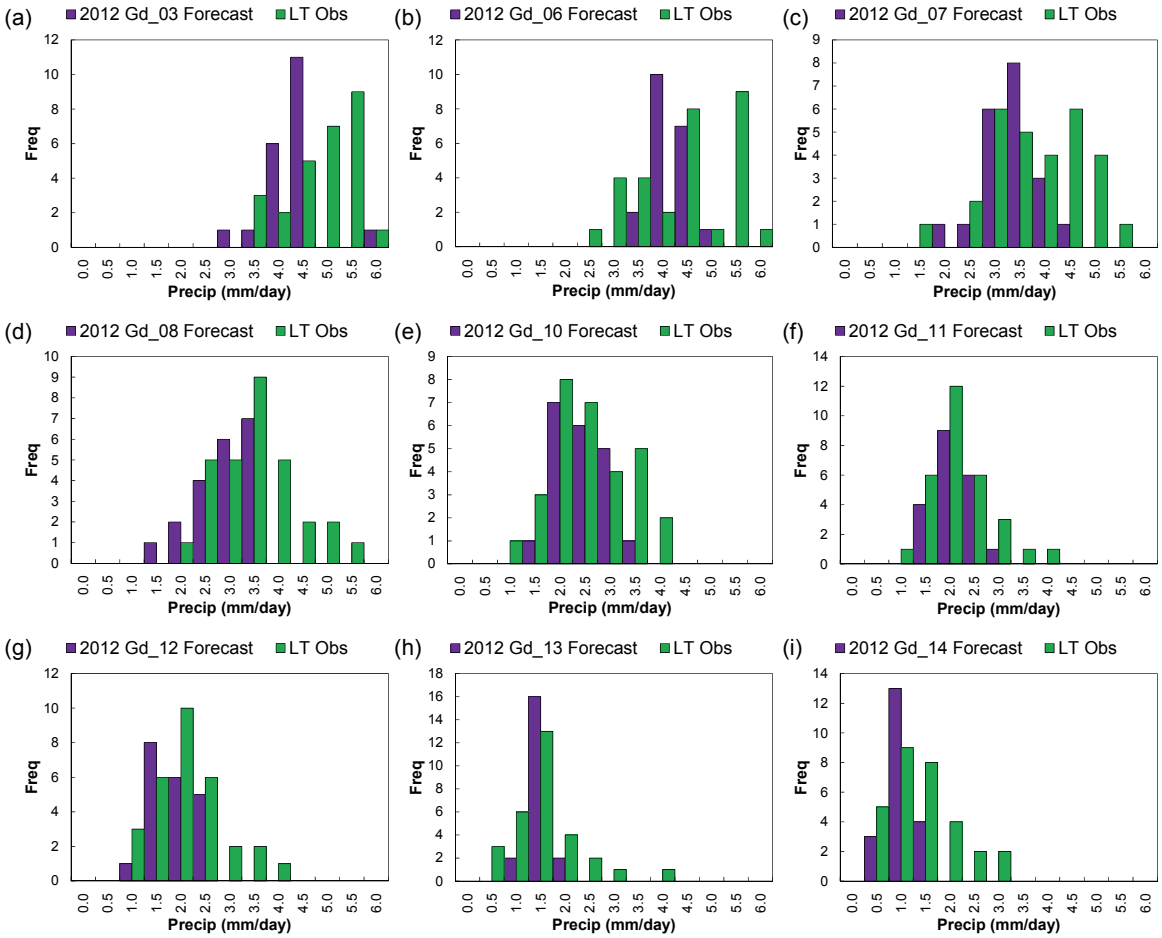



**Figure 5.38:** Histograms of 2012 long-rains forecasts.

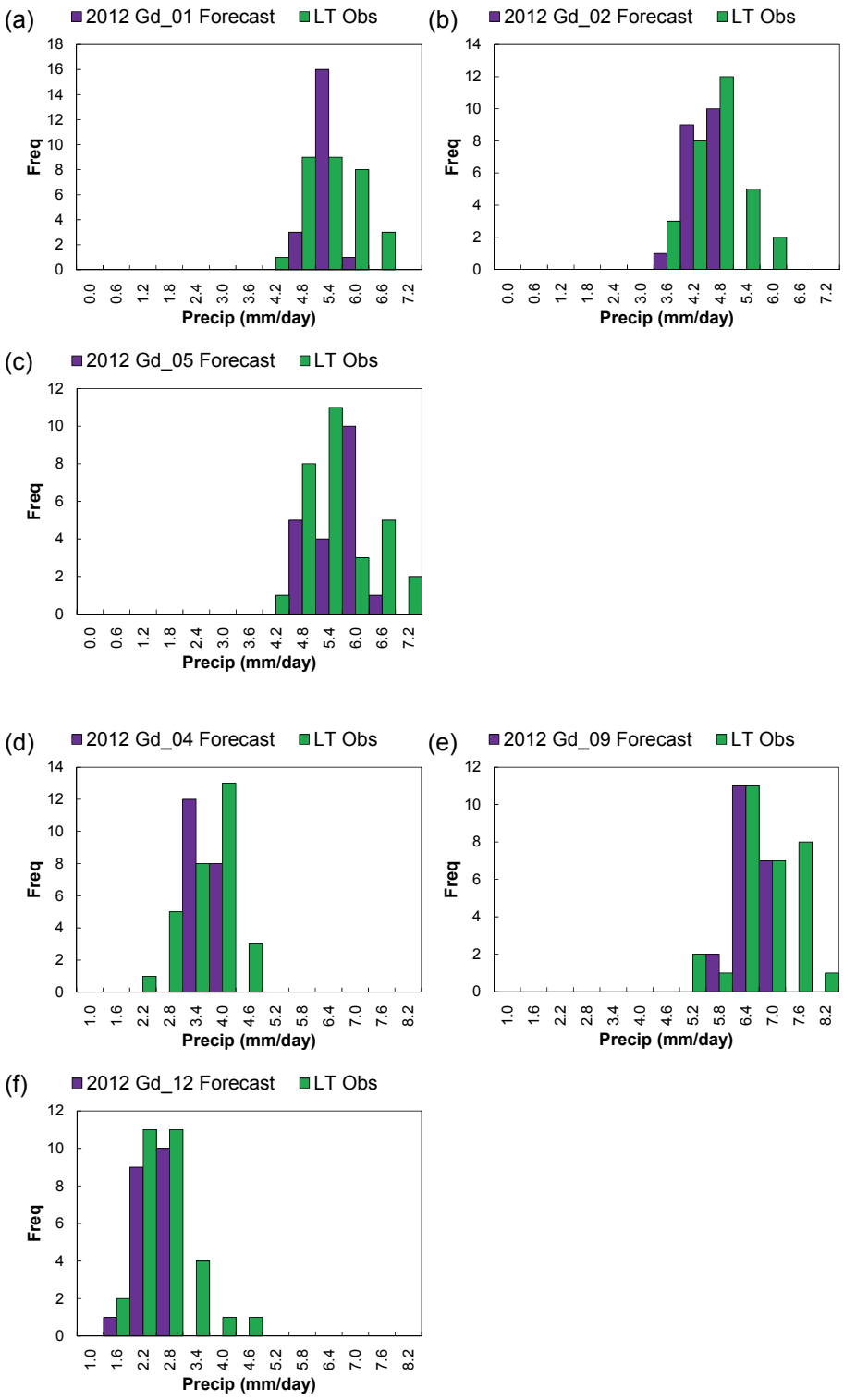



**Figure 5.39:** Histograms of 2012 T-unimodal- (Grids 1, 2, and 5), southern-Sudan- (Grid 4) and Kiremt-rains (Grids 9 and 12) forecasts.

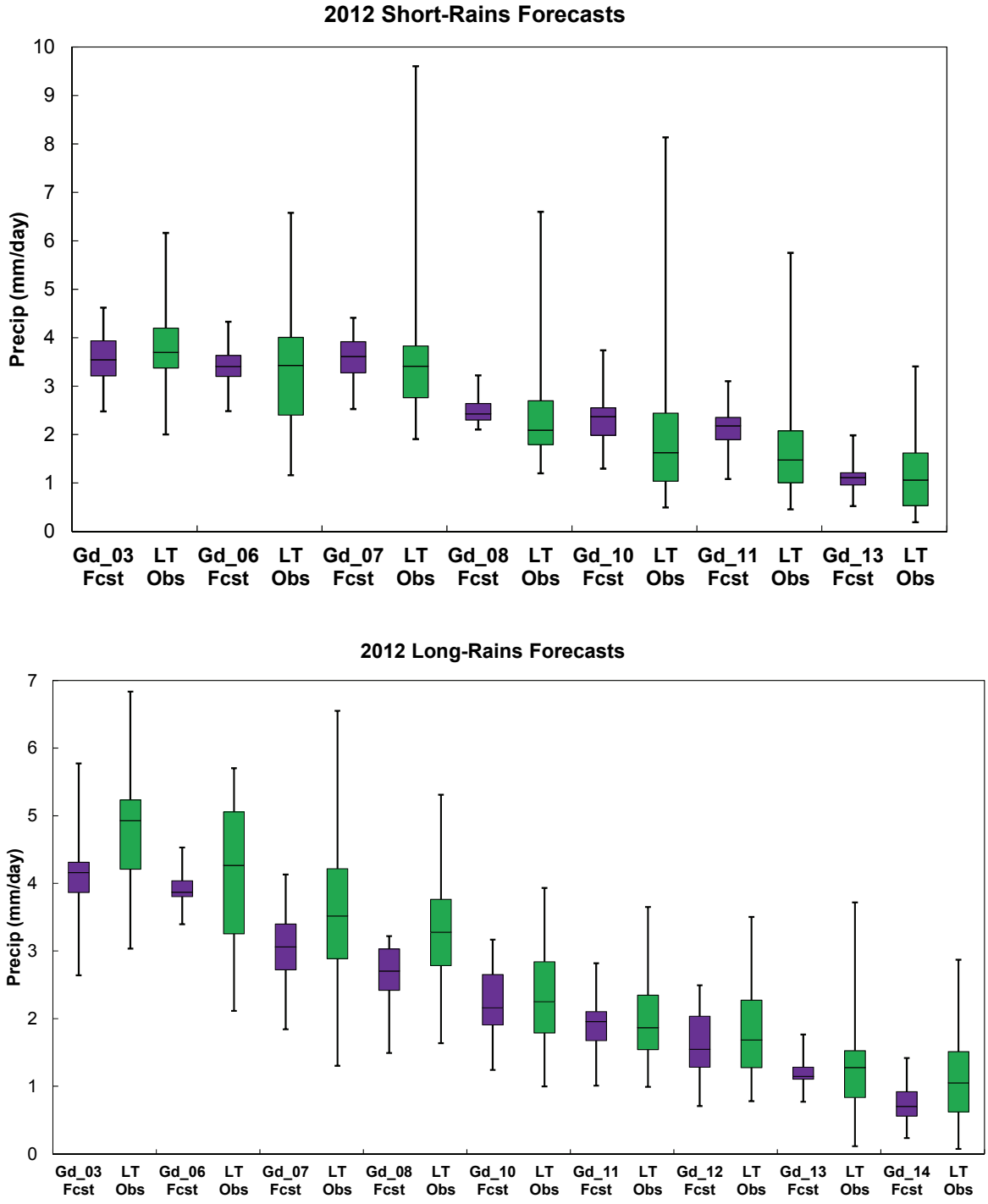
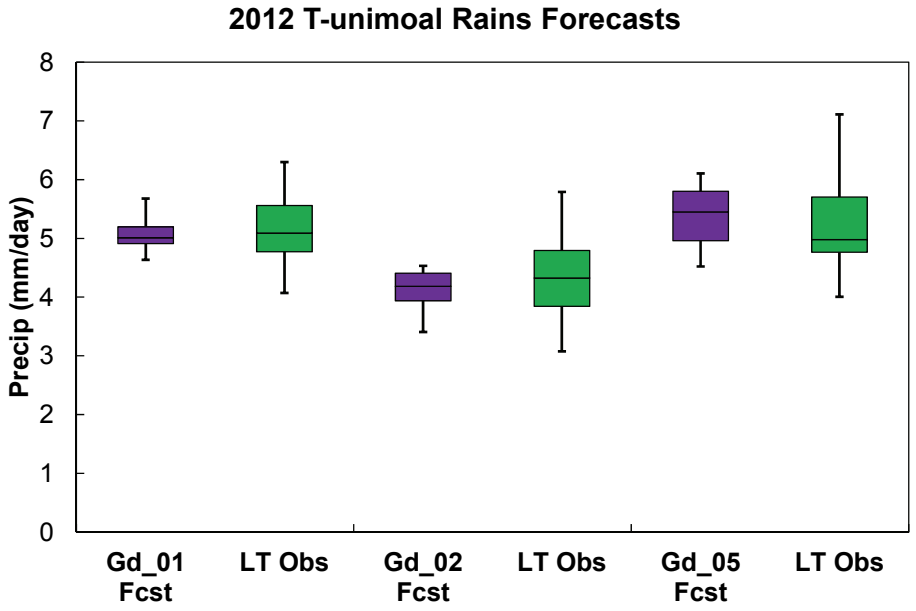



**Figure 5.40:** Boxplots of 2012 short- and long-rains forecasts.



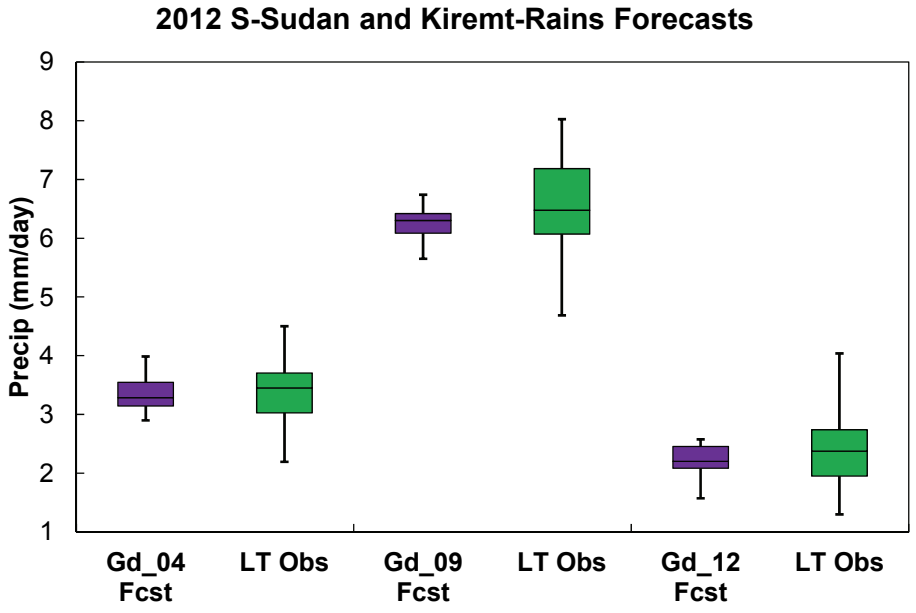




**Figure 5.41:** Boxplots of 2012 T-unimodal-, southern-Sudan-, and Kiremt-rains forecasts.

# CHAPTER 6

# CASE STUDIES IN THE YANGTZE AND CONGO RIVERS

To demonstrate the flexibility of the dipole algorithm, this chapter carries out two additional case studies. In contrast to precipitation predictands adopted in Chapters 4 and 5, streamflow-related data such as flow rate and water level can also be used to drive the dipole algorithm. The direct forecasting of streamflow data can potentially bypass more variable rainfall-runoff processes and avoid the associated uncertainty. Thus, this chapter shows how the dipole algorithm can be applied to the forecasting of the summer Yangtze River flow (Sections 6.1 and 6.2) and the winter Congo River level (Sections 6.3 and 6.4). Section 6.5 performs composite analysis for these two cases to draw possible connections with large-scale variables. Similar to previous two chapters, the last section in this chapter interprets the 2012 forecasting results of the Yangtze River flow and the Congo River level.

## 6.1 Yangtze River Flow

The first predictand examined in this chapter is the inflow data recorded at the Three Gorges Dam, one of the largest hydroelectric dams in the world and located at the midstream of the Yangtze River in China. **Figure 6.1** depicts the geographical location of the Three Gorges Dam as well as its monthly inflow climatology derived from data longer than 100 years. The long-term monthly inflow data presents a single high-flow season occurring during July–September (JAS) with an average flow rate of 30,000 cms.

Because this high-flow season plays a crucial role in determining flood or drought events in this region, the JAS inflow is selected to drive the dipole method.

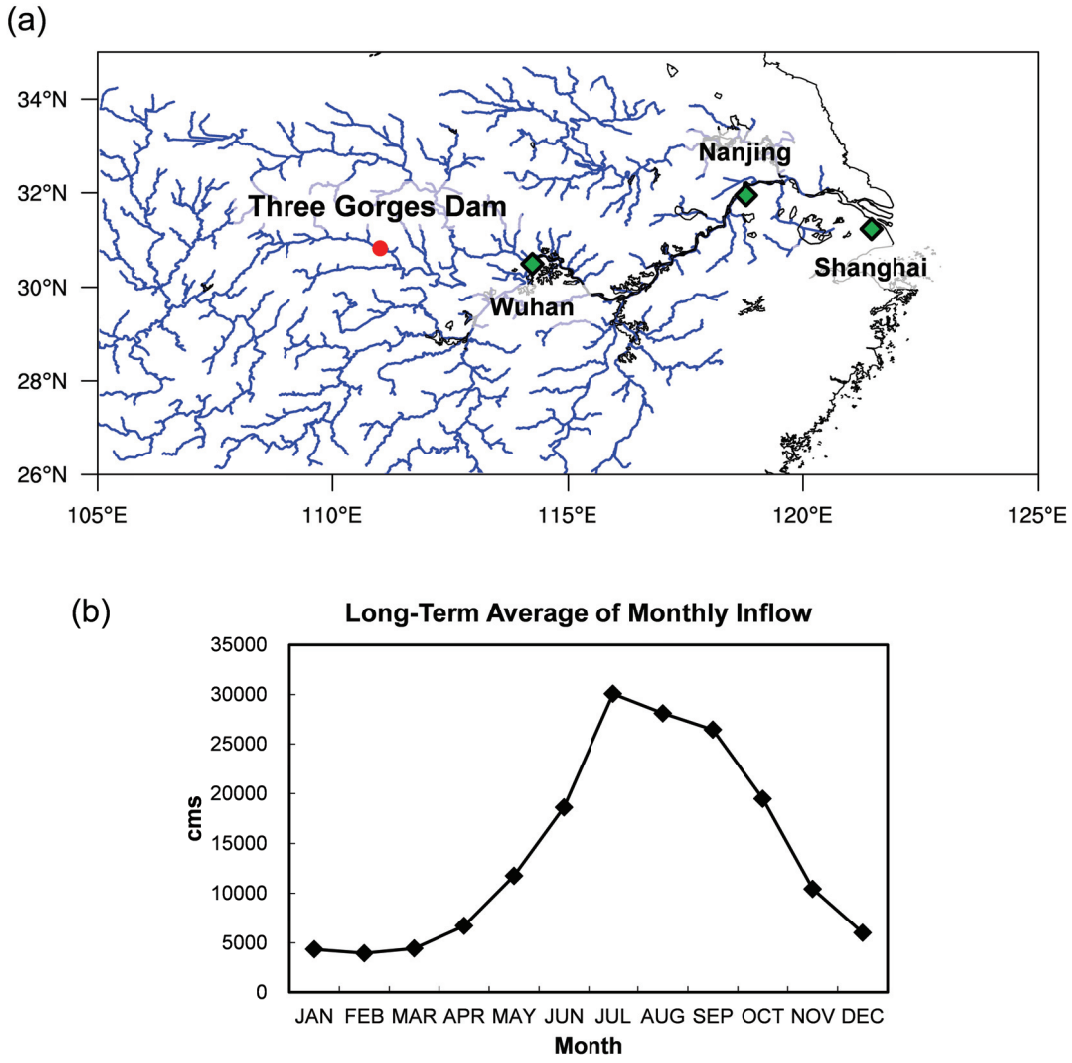



**Figure 6.1:** (a) The location of the Three Gorges Dam at the Yangtze River basin and (b) its associated long-term average (1881–2001) of the monthly inflow.

## 6.2 Dipole Identification for the Yangtze River Flow

In accordance with the description in the last section, the dipole algorithm is applied to the hindcasting of the JAS inflow at the Three Gorges Dam. Although the inflow data is available from 1881 to 2001, to avoid the possible uncertainty of SST data in early periods (lack of observations) but still reserve a minimum data size for statistical analysis, only data in the latest 30 years (i.e., 1972–2001) is adopted as the target external series for forecasting.

The Kaplan SST product is again used for this study as the predictor field. Because precipitation over China is largely modulated by the East Asian Monsoon (Tao and Chen, 1987), an SSTA spatial domain (65°S–65°N, 35°E–285°E) that nearly covers the entire Indian and Pacific Oceans is defined for the dipole identification for the JAS Yangtze River flow (**Figure 6.2**). This domain also includes the key ENSO region (eastern tropical Pacific), which has been shown to influence summer monsoon rainfall as well as streamflow in many studies (e.g., Chang et al., 2000a, b; Wang et al., 2000; Wu et al., 2003; Yang and Lau, 2004; Zhang et al., 2007). Furthermore, to explore the most skillful lead time for forecasting, the dipole algorithm is repeatedly executed from 0- (concurrent) to 12-month lead times. This information and other user-specified parameters used for this case study are summarized in **Table 6.1**.

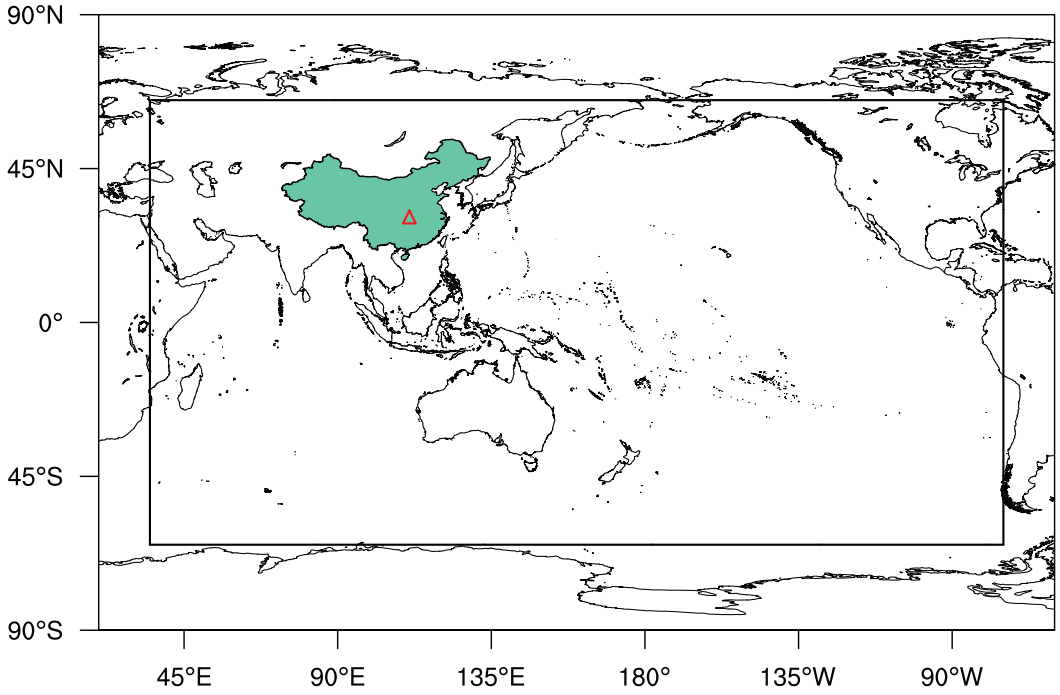



**Figure 6.2:** SSTA predictor domain used for the dipole identification for the Yangtze River flow at the Three Gorges Dam (triangular symbol).

**Table 6.1:** User-specified parameters for the case study in the Yangtze River.

| **Data size** $(h)$ | **Predictor description** | **Predictand description** | **Upper threshold of clustering pixels** $[p \times q]$ | $GSS_{R,\alpha=0.05}$ | $\xi$ | **Lead times** |
|---|---|---|---|---|---|---|
| 30 (1972–2001) | Kaplan SSTA [65°S–65°N, 35°E–285°E] | JAS Inflow at the Three Gorges Dam | $[p = q = 10]$ | 0.25 | 20 | 0, 1, 2, …, 12 (months) |

After the dipole algorithm is applied to the hindcasting of the JAS inflow at the Three Gorges Dam, the subsequent forecasting skills in terms of two scalar accuracy measures (MAE and $\rho$) and four ensemble forecasting measures (Re, LME, UME, and $\gamma$) are recorded in **Table 6.2**. Within the temporal horizon exhibiting forecasting implications (i.e., lead times from three to twelve months), the optimum forecasting skills peak at two different lead times. At short lead times (around 3–4 months), the best CV correlation coefficient is 0.677 (i.e., explaining 46% of the inflow variance), and the lowest UME value of 4394 is observed. At the 8-month lead time, the lowest MAE is 2641 (about 10% of the average inflow), and the highest Re value is 0.833. However, the best LME value appears at the 0-month lead time, indicating that concurrent SSTA conditions can better predict drought events.

**Table 6.2:** Forecasting skills at different lead times for the Yangtze River flow; the best value of each measure is bold and underlined.

| JAS Yangtze River Flow | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 0 | 2802.064 | 0.670 | 0.733 | **-2791.659** | 4724.866 | 0.322 |
| 1 | 2906.288 | 0.600 | 0.800 | -3564.698 | 8529.975 | 0.324 |
| 2 | 2743.233 | 0.602 | 0.767 | -4963.226 | 5560.823 | 0.322 |
| 3 | 2713.755 | **0.677** | 0.700 | -5378.505 | **4393.999** | 0.321 |
| 4 | 2945.631 | 0.643 | 0.800 | -5174.442 | 4884.046 | 0.335 |
| 5 | 3207.988 | 0.508 | 0.667 | -8363.171 | 9438.432 | 0.303 |
| 6 | 3195.248 | 0.491 | 0.633 | -8427.407 | 6399.303 | 0.309 |
| 7 | 2721.613 | 0.594 | 0.800 | -7654.181 | 7662.948 | 0.293 |
| 8 | **2641.365** | 0.614 | **0.833** | -8414.769 | 6359.479 | 0.292 |
| 9 | 2810.717 | 0.606 | 0.767 | -4857.601 | 7819.003 | 0.321 |
| 10 | 2968.560 | 0.465 | 0.633 | -6119.216 | 9501.987 | 0.322 |
| 11 | 3172.067 | 0.416 | 0.733 | -7410.759 | 10219.882 | **0.268** |
| 12 | 2788.386 | 0.562 | 0.767 | -7766.814 | 10698.124 | 0.285 |

To illustrate the correspondence between the ensemble forecasting and observation values, the hindcasting results at two selected lead times (with relatively higher Re values) are plotted in **Figure 6.3**. Over the 30 years, the forecasting envelopes at both lead times can remarkably contain or respond to most of the observations except some extreme events (e.g., in 1994 and 1998). While the 8-month-lead forecasts seem to be able to predict the 1992 drought event, the 4-month-lead forecasts can better respond to the 1998 flood event. Another notable discrepancy between the forecasting traces (at the 4-month lead time) and observations can be seen in 1982, which results from contradictory effects among identified dipoles and will be further examined next.

**Figures 6.4** shows the corresponding dipoles used to generate the hindcasting results at the two selected lead times. At the 8-month lead time, active oceanic areas at the South Indian Ocean and the ENSO region are identified as the major dipoles. At the 4-month lead time, along with the continuously active South Indian Ocean, two other pronounced areas at the East Coast of Africa and the Central Equatorial Pacific are identified and constitute tripole-like patterns. In fact, since the Central to West Equatorial Pacific is usually the complementary ENSO region, the dipoles (and tripoles) at both lead times illustrate the similar significant impacts of ENSO on the JAS inflow at the Three Gorges Dam. Nevertheless, the 1982 discrepancy mentioned earlier is attributed to an inconsistent response from the Central Equatorial Pacific. Unusual positive SSTA extending from the principal ENSO region was observed in this area, but it did not lead to an expected dry event (based on the negative correlation) in the Yangtze River.

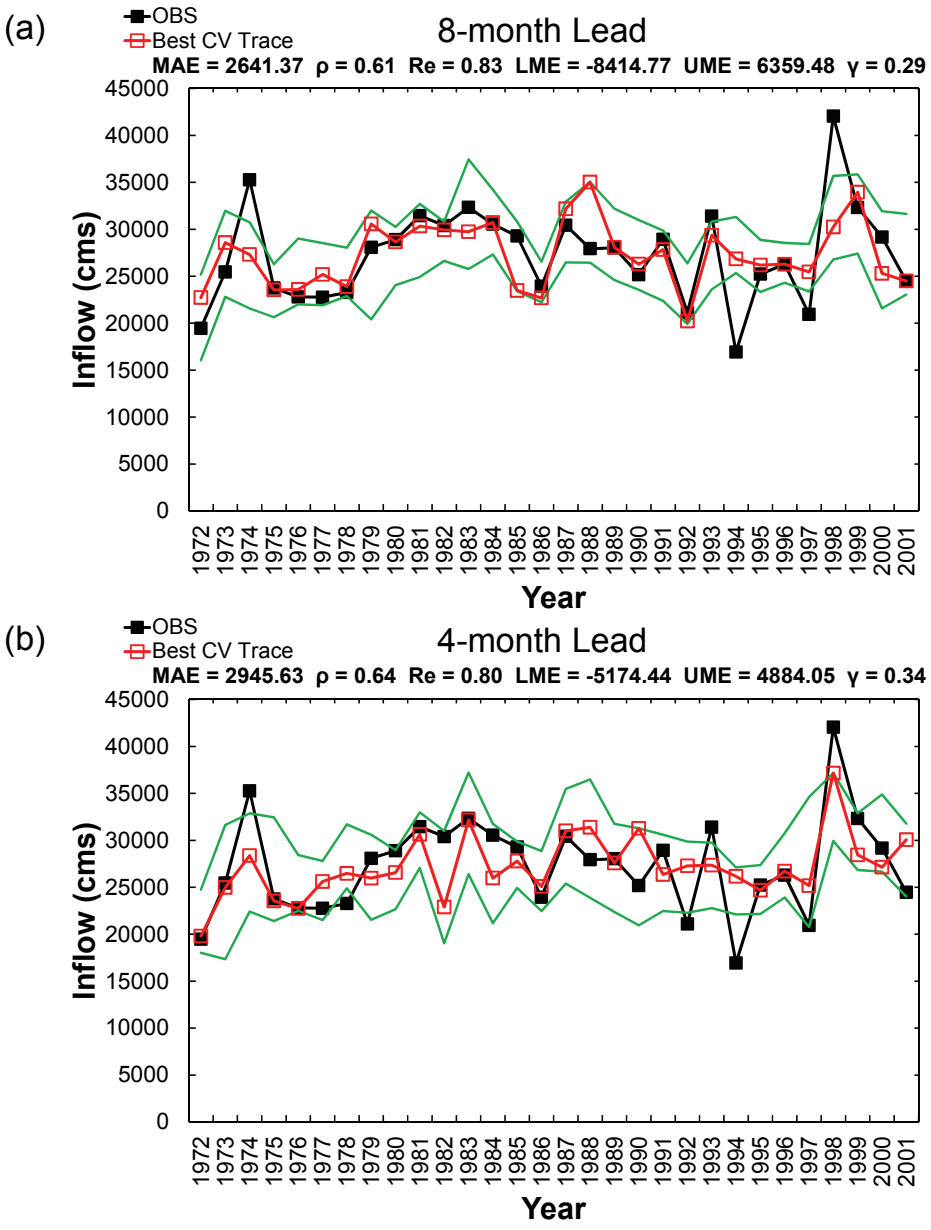



**Figure 6.3:** Optimal hindcasting results at two selected lead times for the Yangtze River flow. The best CV trace is shown in red, and the green envelope is constructed using the maximum and minimum hindcasting values.

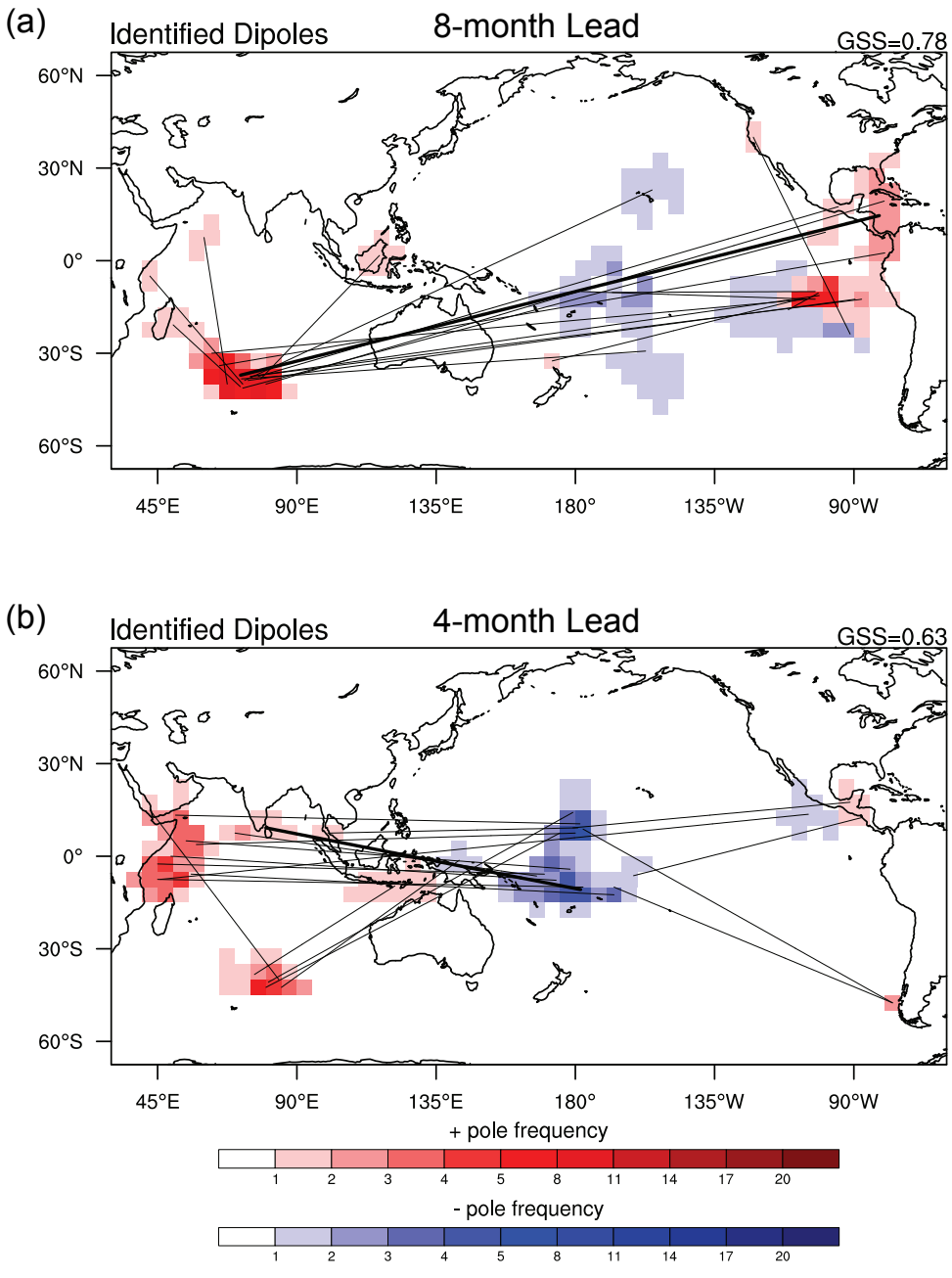



**Figure 6.4:** Corresponding dipoles used to generate the hindcasting results shown in **Figure 6.3**. Thicker and thinner connective lines indicate the orientation of the best and remaining dipoles, respectively. The best GSS among all dipoles obtained during the predictor screening process is denoted at the upper-right corner of each figure.

## 6.3 Congo River Level

The second case study of this chapter is the water-level data recorded at the Kinshasa station, close to the mouth of the Congo River in the Democratic Republic of the Congo. **Figure 6.5** depicts the geographical location of the Kinshasa station as well as its long-term average of monthly water level. Although two peak seasons can be observed, the water level during November–January (NDJ) is more pronounced, almost twice the level during April–June. The NDJ water level is selected to drive the dipole method.

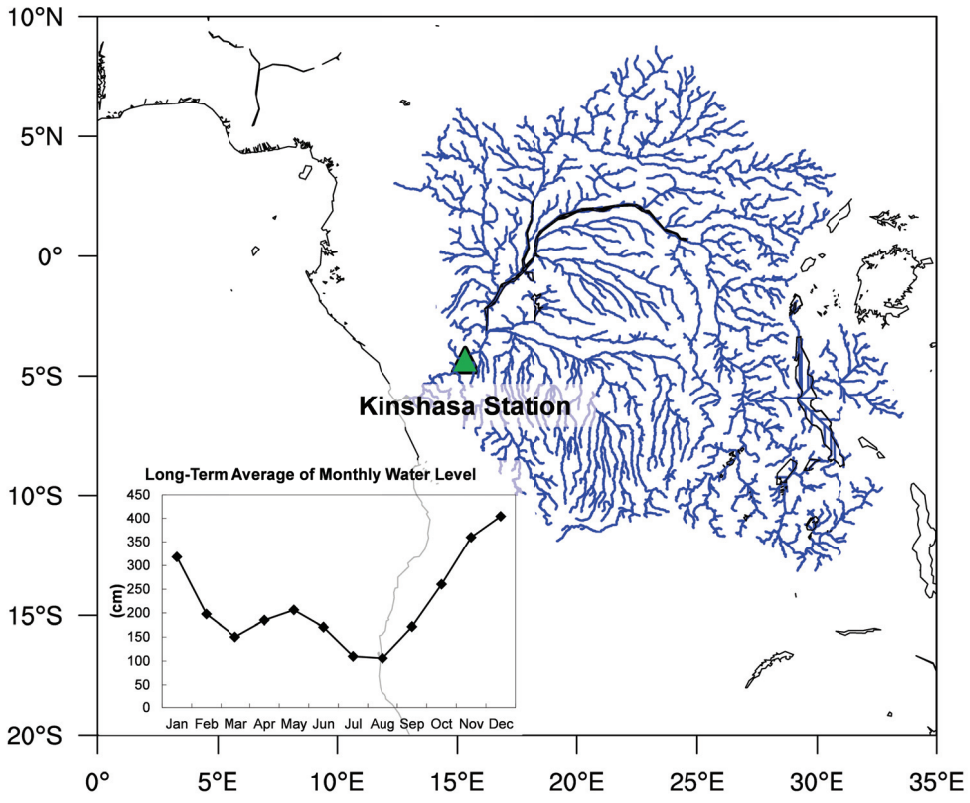



**Figure 6.5:** The location of the Kinshasa station at the Congo River basin and its long-term average (1902–2006) of the monthly water level.

## 6.4 Dipole Identification for the Congo River Level

The dipole algorithm is herein applied to the retrospective forecasting of the NDJ water level at Kinshasa. Although the available water level data is as long as 105 years (1902–2006), for the reasons described in Section 6.2 the most recent 30-year data (i.e., 1977–2006) is adopted as the external series for forecasting.

Literature related to the understanding of large-scale forcing that drives precipitation variability over the Congo River basin is very sparse due to insufficiently accessible data (Samba et al., 2007). Few studies suggest that precipitation over the Congo River basin (or Central to West Equatorial Africa) may be modulated by SSTA from the coasts of Middle to West Africa to the Atlantic Ocean (e.g., Hirst and Hastenrath, 1983; Vizy and Cook, 2001; Todd and Washington, 2004; Balas et al., 2007). Therefore, an SSTA spatial domain (40°S–60°N, 40°W–130°E) that covers large portions the Atlantic and Indian Oceans (**Figure 6.6**) is defined as the input for the dipole algorithm. Besides, a scheme that executes the dipole algorithm from 0- (concurrent) to 12-month lead times is again adopted. This information and other user-specified parameters used for this case study are summarized in **Table 6.3**.

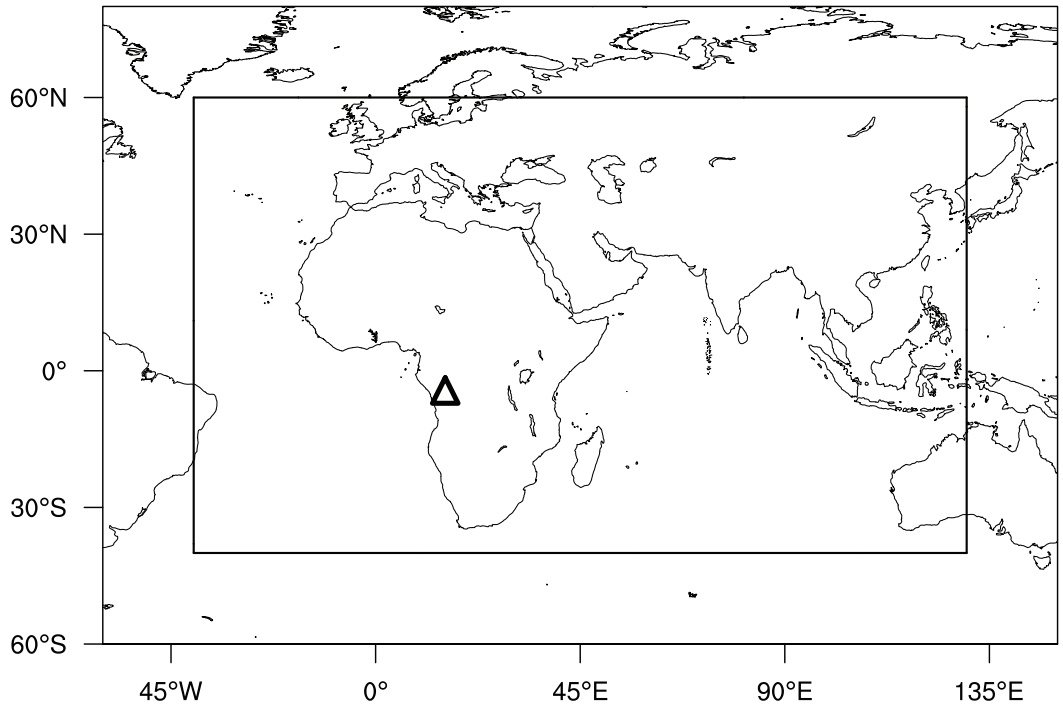



**Figure 6.6:** SSTA predictor domain used for the dipole identification for the Congo River level at the Kinshasa station (triangular symbol).

**Table 6.3:** User-specified parameters for the case study in the Congo River.

| **Data size** $(h)$ | **Predictor description** | **Predictand description** | **Upper threshold of clustering pixels** $[p \times q]$ | $GSS_{R,\alpha=0.05}$ | $\xi$ | **Lead times** |
|---|---|---|---|---|---|---|
| 30 (1977–2006) | Kaplan SSTA [40°S–60°N, 40°W–130°E] | NDJ Kinshasa water level | $[p = q = 10]$ | 0.25 | 20 | 0, 1, 2, …, 12 (months) |

The forecasting skills are recorded in **Table 6.4**. Although considerable skill seem to exhibit multiple peaks at different lead times, the optimum skill obviously appear at the longest lead times (i.e., eleven and twelve months) with lowest MAE 30.24 (about 8% of the average water level), best CV correlation coefficient 0.722 (i.e., explaining 52% of water-level variance), and highest Re value 0.867. Other notable skills (especially when scalar accuracy measures are emphasized) can also be found at the 1- and 4-month lead times with comparable MAE and $\rho$ values.

**Table 6.4:** Forecasting skills at different lead times for the Congo River level; the best value of each measure is bold and underlined.

| NDJ Congo River Level | | | | | | |
|---|---|---|---|---|---|---|
| **LEAD (mths)** | **MAE** | **ρ** | **Re** | **LME** | **UME** | **γ** |
| 0 | 32.226 | 0.607 | 0.567 | -99.693 | 47.687 | 0.340 |
| 1 | 33.336 | 0.610 | 0.733 | -84.528 | 62.000 | 0.330 |
| 2 | 31.241 | 0.599 | 0.700 | -65.596 | 69.768 | 0.354 |
| 3 | 31.803 | 0.560 | 0.633 | -47.837 | 62.393 | 0.336 |
| 4 | 31.281 | 0.673 | 0.667 | -35.827 | 71.651 | 0.372 |
| 5 | 35.798 | 0.381 | 0.567 | **-33.683** | 53.715 | 0.345 |
| 6 | 34.610 | 0.564 | 0.633 | -74.625 | 73.463 | 0.344 |
| 7 | 37.080 | 0.405 | 0.567 | -108.639 | 69.392 | 0.330 |
| 8 | 35.536 | 0.564 | 0.600 | -104.778 | 43.204 | **0.322** |
| 9 | 36.603 | 0.542 | 0.667 | -98.203 | 55.507 | 0.338 |
| 10 | 33.370 | 0.518 | 0.667 | -77.277 | 44.363 | 0.366 |
| 11 | 32.336 | 0.615 | **0.867** | -47.106 | 43.927 | 0.397 |
| 12 | **30.237** | **0.722** | 0.767 | -55.651 | **21.326** | 0.406 |

Due to the multiple forecasting skill peaks, the hindcasting results at three different lead times (with relatively higher Re values and no overlapping periods) are selected and

plotted in **Figure 6.7**. At the 11-month lead time, the forecasting envelope can remarkably respond to most of the observations except two notable events in 1985 and 2003. At the 4-month lead time, whereas the first half of the forecasting envelope can reasonably follow the variations of the observations (even contain the 1985 event), the second half of the envelope appears to lose some of its predictability and generate several false alarms (e.g., events in 1994 and 2000). At the 1-month lead time, both the variability of the forecasting envelope and the false alarm rate appear to improve slightly, but some events in certain years (e.g., 1985 and 2000) are still not depicted well.

**Figures 6.8** shows the corresponding dipoles used to generate the hindcasting results at the three selected lead times. At the 11-month lead time, despite the strong forecasting skills, well-organized dipole patterns barely exist. While three concentrated negative poles are separately centered at the coasts of Angola and Somalia, and the vicinity of Indonesia, positive poles traverse across the Mediterranean Sea, the Arabian Sea, the Bay of Bengal, and all the way to the South China Sea. Although such patterns may be related to the displacement of the ITCZ and African jet streams at long lead times, they may create a too wide forecasting envelope to provide a decisive scenario (i.e., dry or wet). At the 4-month lead time, the dipole patterns seem to be more scattered. Among these scattered patterns, relatively active oceanic areas, including the South Atlantic Ocean, the Mediterranean to the Red Seas, and the South to the East China Seas, can be recognized as the components of major dipoles. In contrast to the patterns delineated above, the dipole patterns identified at the 1-month lead time are very organized. The zonal-oriented dipoles that comprise negative poles located at the South Atlantic and

positive poles located at the Central Indian Ocean illustrate a clear east-west temperature gradient. This may dominate westerly anomalies crucial to the precipitation over the Congo River basin.

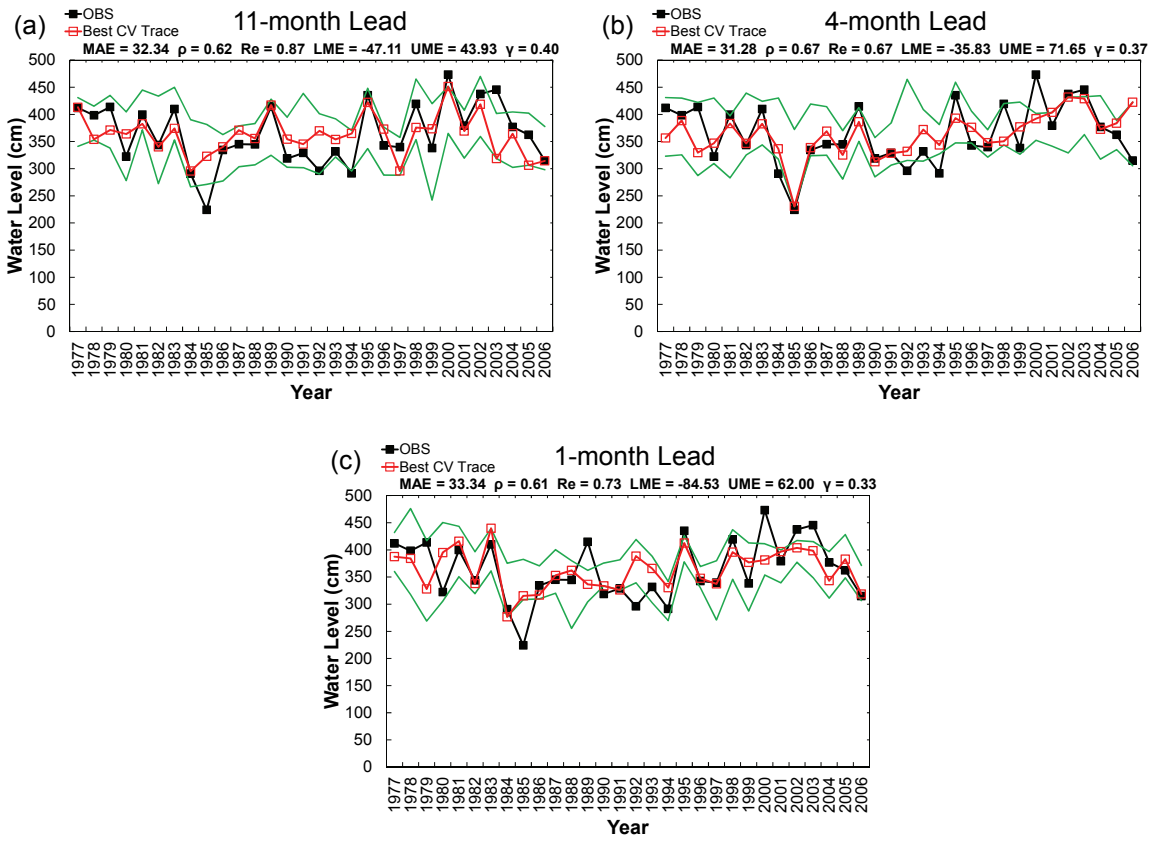



**Figure 6.7:** Optimal hindcasting results at three selected lead times for the Congo River level. The best CV trace is shown in red, and the green envelope is constructed using the maximum and minimum hindcasting values.

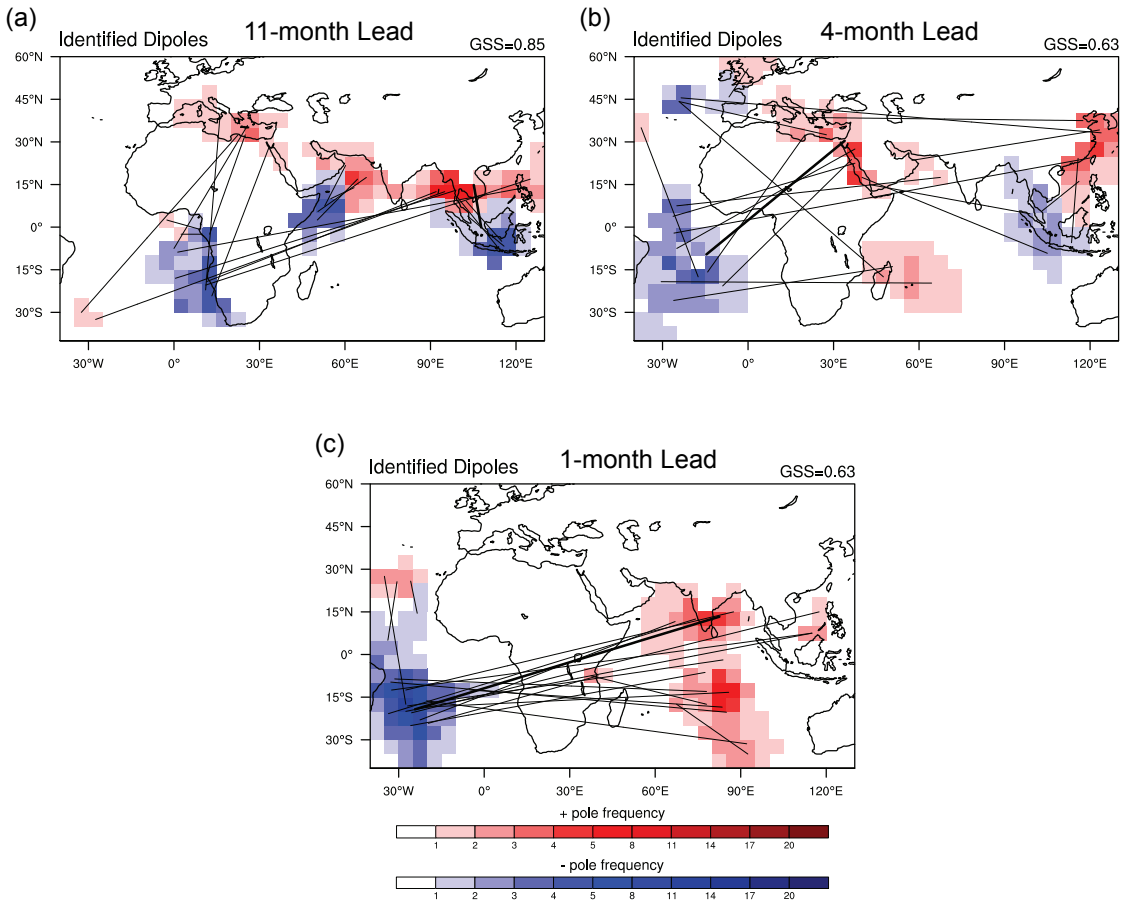



**Figure 6.8:** Corresponding dipoles used to generate the hindcasting results shown in **Figure 6.7**. Thicker and thinner connective lines indicate the orientation of the best and remaining dipoles, respectively. The best GSS among all dipoles obtained during the predictor screening process is denoted at the upper-right corner of each figure.

## 6.5 Diagnostic Analysis of Large-Scale Variables

As conducted in previous two chapters, diagnostic analysis of large-scale variables pertaining to wet or dry composite years can facilitate the understanding of hypothetical mechanisms that dominate streamflow-related variables at the Yangtze and Congo Rivers. Again, to generate composite maps, data at 500mb (geopotential height and vector wind

anomalies) and surface (SLP and vector wind anomalies) levels retrieved from the NCEP/NCAR Reanalysis dataset (Kalnay et al., 1996) are used.

According to the time series of the Yangtze River flow or the Congo River level, ten wet and ten dry years (out of 30 years) when the river flow or the water level is higher and lower than its long-term 67th and 33rd percentiles are identified and listed in **Table 6.5**. Notably, for both the Yangtze and Congo Rivers, consecutive wet or dry events can be observed. For instance, for the Yangtze River, high-flow events occurred from 1981 to 1985, and low-flow events occurred from 1975 to 1978. For the Congo River, a low water level was consecutively recorded from 1984 to 1986 and from 1990 to 1994.

**Table 6.5:** Wet and dry years (corresponding to the Yangtze River flow and the Congo River level) used in composite analysis.

| Region | Season | Composite Years | |
|---|---|---|---|
| | | Wet | Dry |
| Yangtze River | JAS | 1974, 1981, 1982, 1983, 1984, 1985, 1987, 1993, 1998, 1999 | 1972, 1975, 1976, 1977, 1978, 1986, 1992, 1994, 1997, 2001 |
| Congo River | NDJ | 1977, 1979, 1981, 1983, 1989, 1995, 1998, 2000, 2002, 2003 | 1980, 1984, 1985, 1986, 1990, 1991, 1992, 1993, 1994, 2006 |

*6.5.1 Yangtze River Flow*

**Figure 6.9** illustrates composite maps in the wet and dry JAS years of the Yangtze River. In general, the composite maps have difficulty in revealing clear out-of-phase patterns. In other words, mechanisms that drive wet conditions do not necessarily reverse to drive

dry conditions over the Yangtze River basin, indicating the complicated impacts of large-scale patterns on the streamflow at the Yangtze River. Regardless of such complicated relationships, several more pronounced patterns are summarized as follows.

At the 500mb level during the wet years, cyclonic anomalies over East Asia that enhance the southwestern monsoon intruding into southern and southeastern China can be found. Meanwhile, in the North Pacific, high anomalies centered at the Gulf of Alaska are surrounded by extensive low anomalies. This may be conducive to route Pacific jet streams thereby playing an indirect role in remotely influencing the moisture transport over the Yangtze River basin. Besides, equatorial easterly anomalies in the eastern tropical Pacific related to La Niña conditions can also be observed during the wet years. On the other hand, during the dry years, enhanced and extensive low anomalies occupy high latitude regions and the western subtropical Pacific, which induce easterly anomalies in southern China. In the North and East Pacific areas, opposite patterns such as low anomalies at the Gulf of Alaska and equatorial westerly anomalies can be found. Further, in the Southern Hemisphere, extensive low anomalies and significant high anomalies emerge from the subtropical and high latitude areas, respectively.

At the surface level, most of the aforementioned patterns (e.g., cyclonic anomalies over East Asia during the wet years, and the opposite pressure patterns at the Gulf of Alaska) seem to remain but with weakened intensity except for some patterns in the Southern Hemisphere. In both the wet and dry years, elongated high anomalies appear at mid-to-

high latitude areas (much more enhanced in dry years). However, how such anomalies affect the precipitation and streamflow at the Yangtze River requires further investigation.

To further examine the plausibility of the identified dipoles and to elucidate the oceanic response, wet-minus-dry SSTA composite maps with respect to the Yangtze River flow at different lead times are generated in **Figure 6.10**. The figure indicates that high (low) JAS inflow at the Three Gorges Dam attributed to anomalous monsoon rainfall typically occurs in the decaying years of El Niño (La Niña) events, which is consistent with results shown in previous studies (e.g., Chang et al., 2000a, b). The correspondence between the components of dipoles and the prominent oceanic areas in the composite maps is remarkable. For instance, the South Indian Ocean and the East Coast of Africa at the 4-month lead time and the key ENSO region at the 8-month lead time are all successfully identified by the algorithm (see **Figure 6.4**). Furthermore, the algorithm can also identify the warming signals over the eastern Indian Ocean at short lead times (not shown), which have been noted in several studies (e.g., Yang and Lau, 2004; Xu et al., 2007).

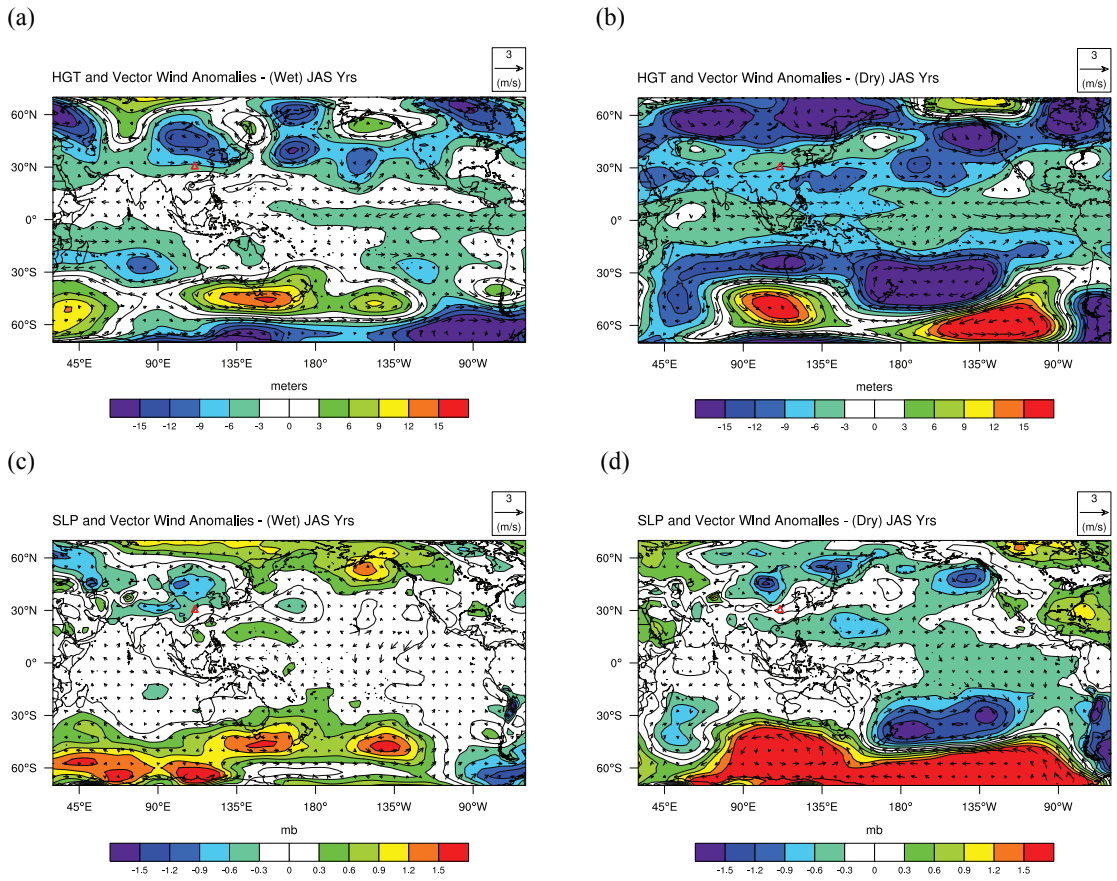



**Figure 6.9:** JAS composite maps of NCEP/NCAR Reanalysis data for wet [(a) and (c)] and dry [(b) and (d)] years of the Yangtze River. The upper panel is the 500mb geopotential height and vector wind anomalies and the lower panel is the SLP and vector wind anomalies.

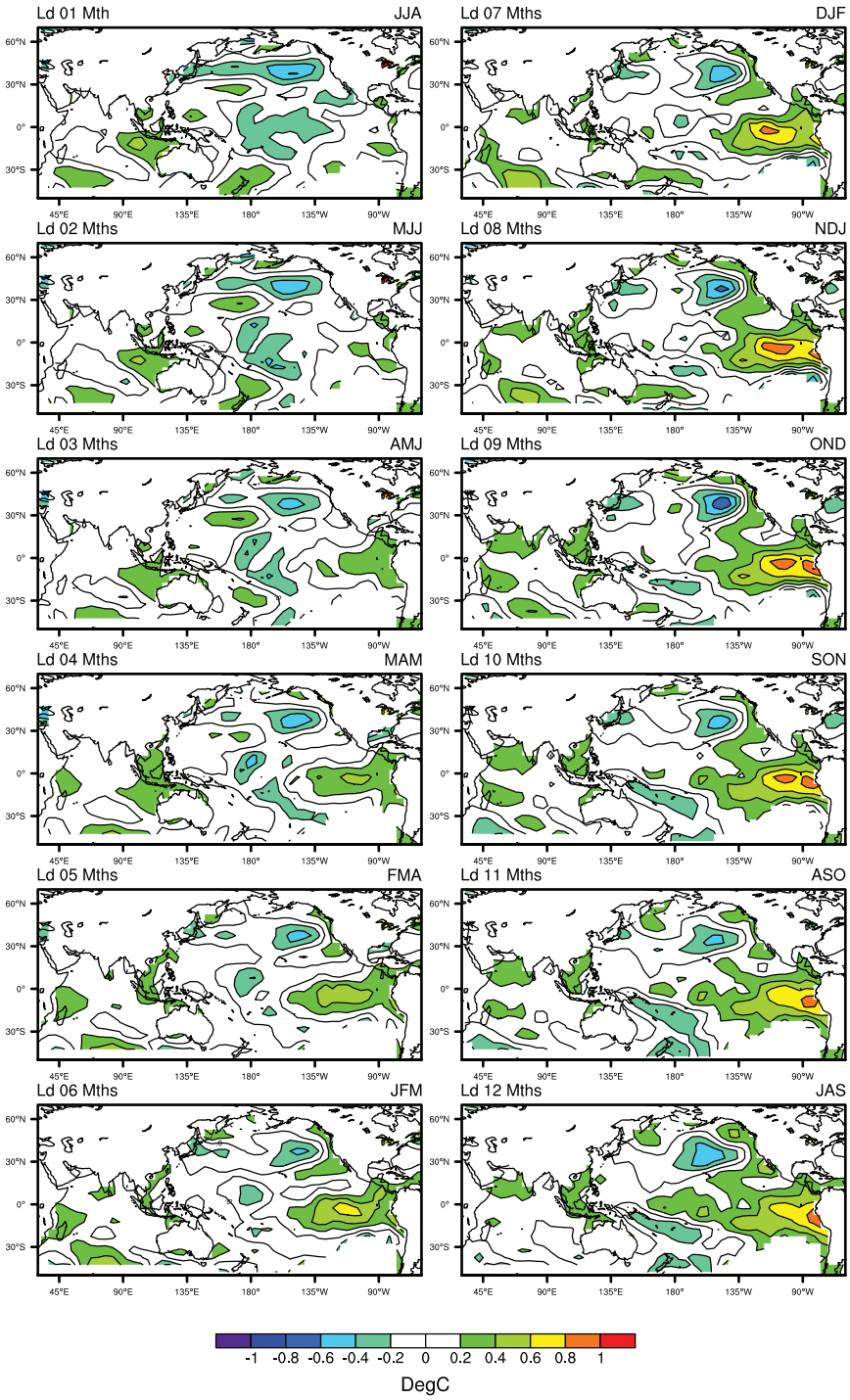



**Figure 6.10:** Evolution of SSTA composites in wet-minus-dry (JAS) years of the Yangtze River.

*6.5.2 Congo River Level*

**Figure 6.11** illustrates composite maps in the wet and dry NDJ years of the Congo River. Similar to the composite maps of the Yangtze River, clear out-of-phase patterns can hardly be revealed. More pronounced anomalies addressed as follows seem to appear at the tropical areas and the Northern Hemisphere. At the 500mb level during the wet years, equatorial westerly and easterly anomalies can be found in the Atlantic and Indian Oceans, respectively. Meanwhile, high anomalies centered at the Coast of Western Europe and low anomalies located at Eastern Europe form a dipole-like pattern. On the other hand, during the dry years, easterly anomalies observed over Tanzania and southern Democratic Republic of the Congo may impede the incursion of moist air from the Atlantic Ocean. Instead of a reversed dipole pattern, only low anomalies located at West Africa appear to be more significant in the Northern Hemisphere.

At the surface level, except that the equatorial easterly anomalies in the Indian Ocean and the high anomalies at the Coast of Western Europe remain significant, most of the aforementioned patterns become weakened or even disappear. Nevertheless, because of lack of related studies that support these patterns, the underlying mechanisms that drive the rainfall and streamflow variations over the Congo River basin are still inconclusive.

**Figure 6.12** shows wet-minus-dry SSTA composite maps with respect to the Congo River level at different lead times. If the emphasis is on predefined SSTA domain used for the dipole identification, the remarkable correspondence between the dipole components and the active oceanic areas can be seen. For instance, the coast of Angola

at the 11-month lead time, the Mediterranean Sea at the 4-month lead time, and the Central Indian Ocean at the 1-month lead time are all successfully identified by the algorithm (see **Figure 6.8**). Moreover, the figure indicates that high NDJ water level at the Kinshasa station typically occurs when ENSO experiences a transition from a La Niña to El Niño phase, and vice versa. Even though the weak relationship between the rainfall in the Congo River basin (or West Central Africa) and ENSO has been addressed in previous studies (e.g., Amarasekera et al., 1997; Balas et al., 2007), the evolution of SSTA composites that clearly depicts the ENSO transition is first carried out in this study.

Since the ENSO transition may influence the rainfall and streamflow variations over the Congo River basin, several strategies can be employed to improve the forecasting skill. For example, a larger SSTA domain that covers the eastern tropical Pacific can be applied to the dipole algorithm to identify if significant dipole predictors reside in the ENSO region. In addition, the continuous monitoring of ENSO conditions can also assist in forecasting the onward scenario of the Congo River level.

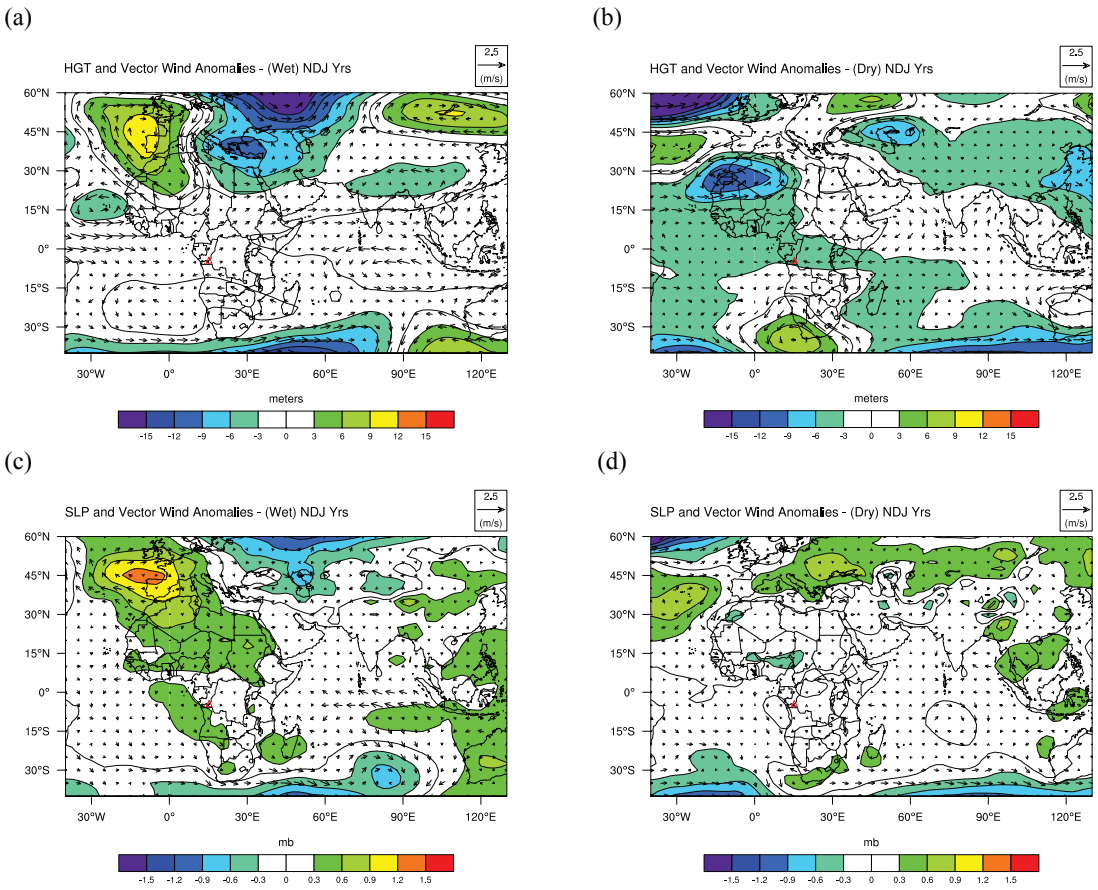



**Figure 6.11:** As in **Figure 6.9**, but for NDJ composite maps of the Congo River.

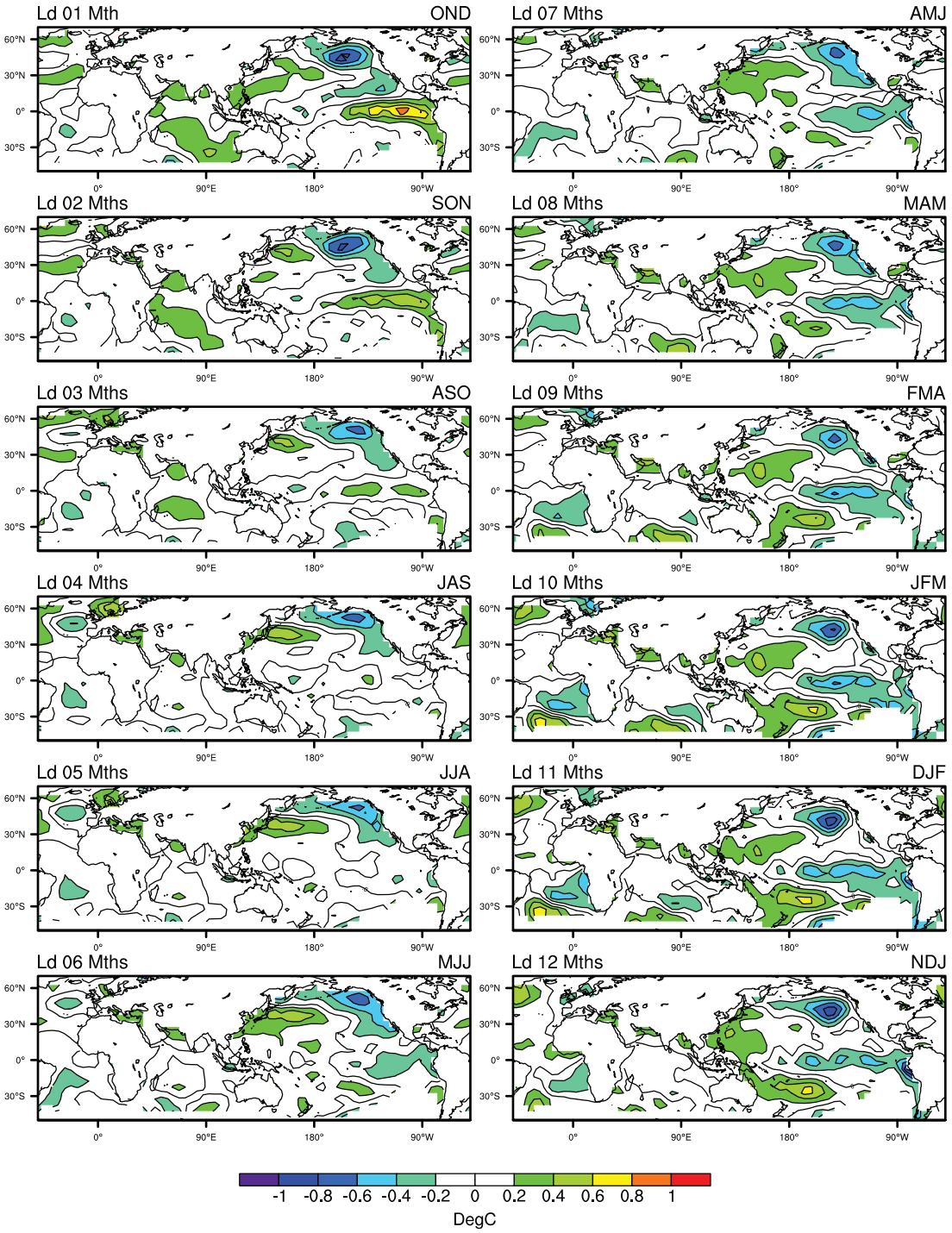



**Figure 6.12:** Evolution of SSTA composites in wet-minus-dry (NDJ) years of the Congo River.

## 6.6 Demonstration of 2012 Forecasts

An operational forecasting framework for both the Yangtze River flow and the Congo River level is carried out in this section. Similar to the framework introduced in previous two chapters, the dipole predictors at lead times pertaining to the highest Re values for each predictand are selected to prepare the 2012 forecasts with the newest available SSTA data (as recent as January 2012). Because the SSTA data at the optimal lead times for the forecasting of both the 2012 JAS inflow at the Three Gorges Dam and NDJ water level at the Kinshasa station are available, forecasting results presented next do carry with the highest Re values and require no further updates.

**Table 6.6** summarizes the key statistics of the 2012 forecasts and the associated long-term climatological values. To reveal the characteristics of forecasting distribution, histograms and boxplots of the forecasting results are also plotted in **Figures 6.13** and **6.14**, respectively.

The 2012 forecast of the JAS inflow at the Three Gorges Dam indicates a below-normal flow rate, partially attributed to the La Niña signals observed in early seasons. However, because only 20% of the ensemble forecasting traces are lower than the 33rd long-term percentile (indicating a dry scenario), the expected dry condition may be moderate. The forecasting distribution seen from the histogram and boxplot appears to be quite concentrated and symmetric. In light of such distribution and the high Re value (0.83), this forecast is characterized by a high confidence level.

The 2012 forecast of the NDJ water level at the Kinshasa station presents a slightly above-normal scenario. However, the forecasting distribution seen from the histogram and boxplot appears to have a wide spread. Although 30% of the ensemble forecasting traces are higher than the 67th long-term percentile (i.e., wet), 20% of them are lower than the 33rd long-term percentile (i.e., dry). This implies the climatic background is somewhat mixed. In accordance with the newest precipitation observation over the Congo River basin obtained from the Global Precipitation Climatology Centre website (ftp://ftp-anon.dwd.de/pub/data/gpcc/html/download_gate.html), the true condition of the water level tends to be near or above normal.

**Table 6.6:** Summary of 2012 inflow and water-level forecasts in the Yangtze and Congo Rivers.

| **Statistics** | **Predictand** | |
|---|---|---|
| | Yangtze Inflow | Congo River Level |
| Max | 29606.34 | 436.58 |
| Mean | 26432.53 | 372.57 |
| Min | 23564.18 | 299.00 |
| Re | 0.83 | 0.87 |
| LT Max | 42051.62 | 473.33 |
| LT Mean | 27270.57 | 362.83 |
| LT Min | 16924.75 | 224.00 |

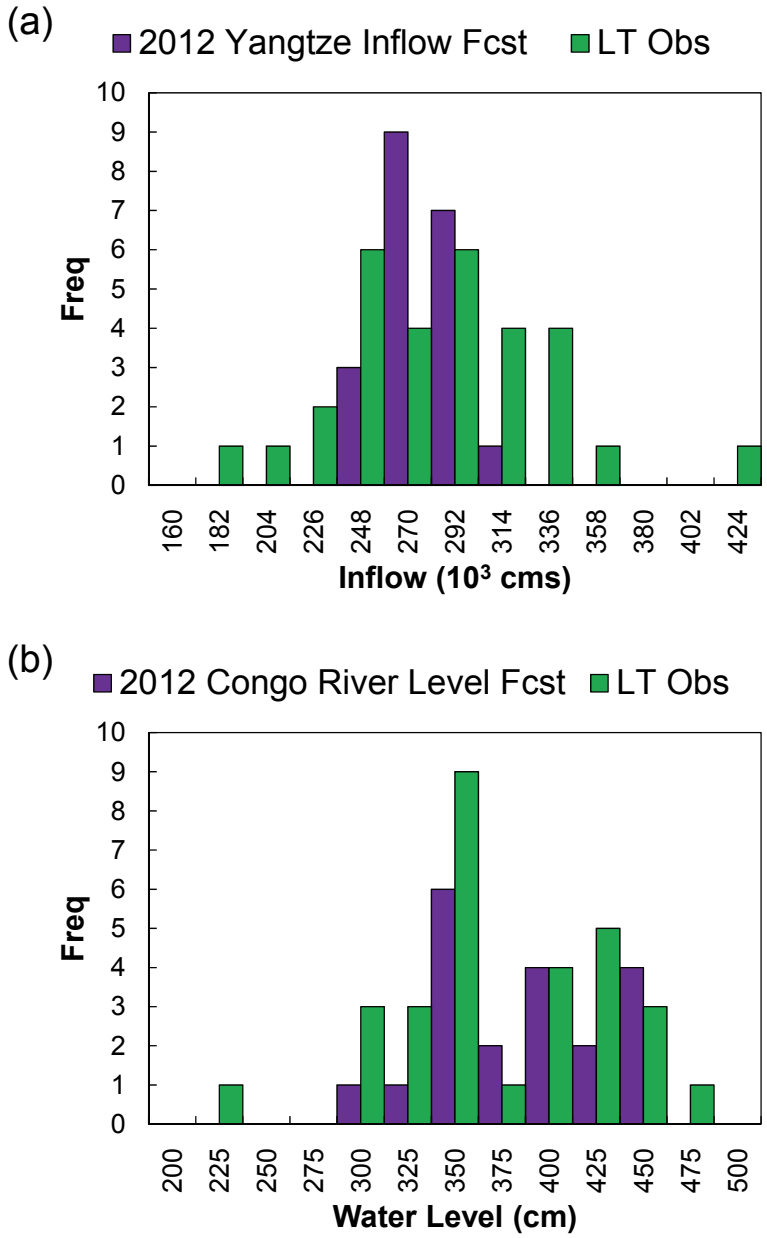



**Figure 6.13:** Histograms of 2012 forecasts of (a) the Yangtze River flow and (b) the Congo River level.

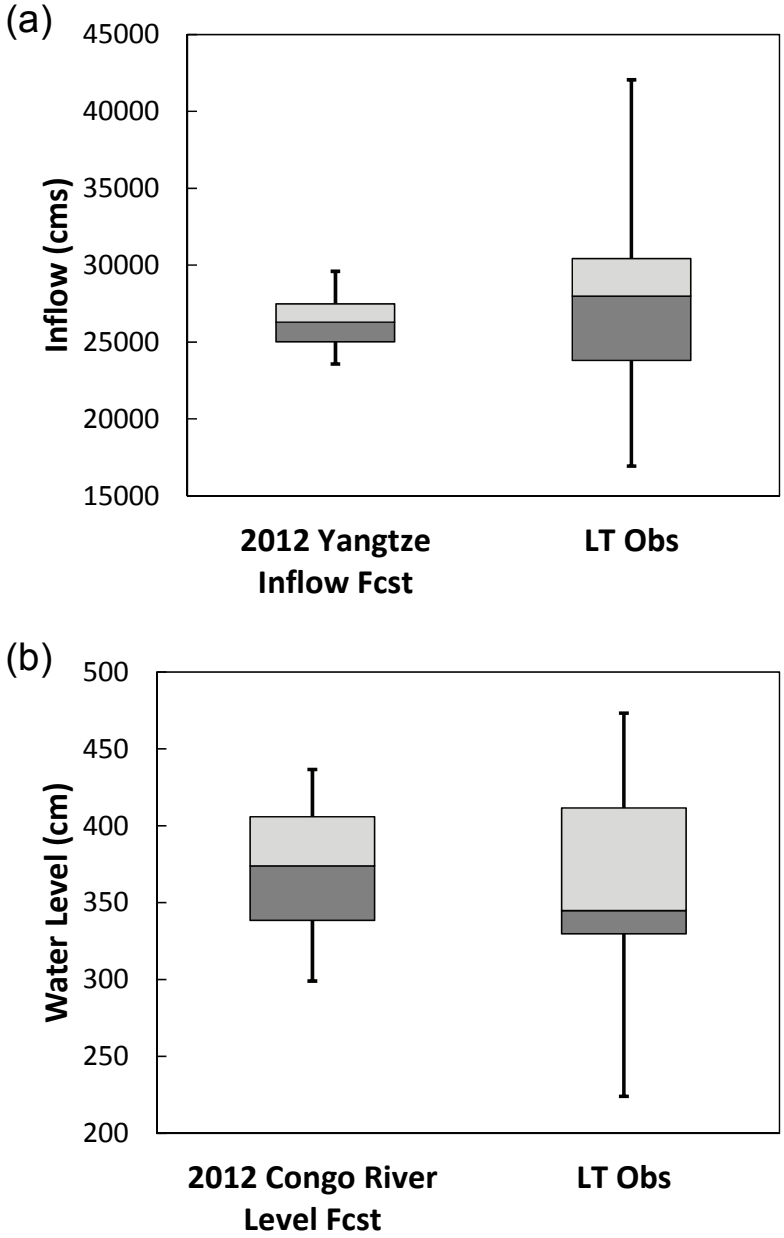



**Figure 6.14:** Boxplots of 2012 forecasts of (a) the Yangtze River flow and (b) the Congo River level.

# CHAPTER 7

# CONCLUSIONS AND RECOMMENDATIONS

## 7.1 Research Contributions and Findings

This dissertation introduced a new seasonal forecasting method for hydro-climatic variables based on SSTA dipoles. The dipole teleconnection method was demonstrated through case studies in the Southeast US, East and Central Africa, and China and was shown to generate reliable multi-season forecasts. The method compared favorably with other existing methods, including CCA, ENSO (or IOD) regression, and climatological forecasting. Furthermore, the method is flexible to forecast various hydro-climatic variables, including precipitation, temperature, winds, and streamflow, among others.

The dipole method was applied to the hindcasting of various predictands, including seasonal precipitation over the ACF basin, different rain types in East Africa, and streamflow-related variables in the Yangtze and Congo Rivers. Specific research findings from these case studies are summarized below:

*Southeast US – ACF:*

- The case study in the ACF basin in the southeastern US shows that DJF precipitation for both UACF and LACF can be predicted with high forecasting skill; the ensemble forecasting envelope contains the true observations with high reliability (~ 0.833), and the best dipole trace can explain more than 50% of precipitation variance with a lead time of at least three months.

- Although forecasting skill decreases in SON, JJA, and MAM, forecast reliability is at least 0.600, and the best dipole trace can still explain a minimum of 30% of precipitation variance. Besides, if events impacted by prominent hurricanes/tropical storms are removed, the updated forecasting skills in JJA and SON improve substantially.
- The dipole patterns identified for DJF precipitation in both UACF and LACF present some connections to ENSO, but the predictor-predictand relationship can further be enhanced by the dipole structure and the auxiliary signals from the North Atlantic Ocean (near the coast of Iceland).
- Because precipitation in other seasons are governed by internal atmospheric variability and are affected by nonlinear tropical disturbances, the identified dipole patterns are scattered. However, several active oceanic areas, such as the North Pacific and Atlantic Oceans and the West Coast of North America, can still be identified as dipole components conducive to seasonal precipitation.
- Composite analysis of the ACF case study reveals that high (low) seasonal precipitation in the southeastern US is universally dominated by negative (positive) geopotential height anomalies over the Central US. This accompanies southwesterly (northeasterly) anomalies that favor (impede) moisture transport from the Gulf of Mexico. Meanwhile, some extratropical patterns that may influence jet stream location are also noticed and deserve further attention.

*East Africa:*

- The case study in East Africa shows that the short rains in Grids 11 (over the border between Ethiopia, Somalia, and Kenya) and 3 (over Uganda and Lake Victoria) can be predicted with strong forecasting skill. The forecast reliability is 0.900 and 0.733, and the best dipole traces can explain more than 50% of precipitation variance in the 1-month lead time.
- The best lead time for predicting the East African short rains is consistently found to be one month for every grid, implying the eminent concurrent interactions between the short rains and SST forcing.
- Some of the dipole patterns identified for the short rains are zonal-oriented and resemble the IOD in the 1-month lead time, while other patterns emphasize the southwestern Indian Ocean in longer lead times.
- The East African long rains in Grids 3 and 7 (western Kenya) can be predicted with strong forecasting skill (i.e, with reliability of 0.800 and 0.767, respectively), and the best dipole traces can explain more than 50% of precipitation variance in the 11-month lead time.
- The long rains in some grids exhibit a prominent downward trend over the latest 30 years. However, such trend can still be depicted by the identified dipoles and associated forecasting envelopes.
- The dipole patterns identified for the long rains are mixed with multiple active oceanic areas, including the South Atlantic Ocean, the Mediterranean Sea, the southwestern Indian Ocean, and the Arabian Sea, varying from grid to grid, but some coherence between patterns in adjacent grids is discernible.

- The Tanzania unimodal rains in delineated Grid 2 (the mountainous region in northwestern Tanzania) can be predicted with strong forecasting skill. The forecast reliability value is 0.767, and the best dipole trace can explain more than 50% of precipitation variance in long lead times.
- Common dipole patterns composed of a positive pole at the southwestern Indian Ocean and a negative pole at the subtropical North Atlantic can be identified for the Tanzania unimodal rains in different grids.
- The southern Sudan and the Kiremt rains can only be predicted with modest forecasting skill (i.e., with reliability of 0.667), and the best dipoles traces can explain about 32–44% of precipitation variance in median to long lead times. As a result, the identified dipole patterns are relatively scattered.
- Applying a larger SSTA domain that covers the Pacific Ocean to the dipole identifications for selected grids can further improve the forecasting skills. More importantly, ENSO-related and other eminent signals over the Pacific Ocean are proven to be conducive to the forecasting of most of the rain types in East Africa.
- Composite maps in wet and dry short-rains years for Grids 3 and 11 reproduce the Walker-like circulation patterns over the Indian Ocean. The comparison between the identified dipole patterns and the SSTA composite maps clarifies that although ENSO is generally correlated with the short rains, its key region (i.e., the eastern tropical Pacific) exhibits a weak connection to the short rains.
- Composite maps regarding the long rains in Grids 3 and 11 illustrate notable westerly (easterly) anomalies over Central Africa that carry (impede) more unstable, moist air from the Congo Basin and the Gulf of Guinea during the wet

(dry) years. Besides, the SSTA composites reveal weak ENSO signals in long lead times. However, some other extratropical patterns contradict those found in previous studies that used data in earlier decades, implying some general change in large-scale circulation after the late '90s.

- Composite maps at the 500mb and surface levels regarding the Kiremt rains in Gird 9 barely show any significant patterns over tropical regions, which may also be attributed to the change in large-scale circulation in the recent decade. Beyond the tropical regions, Antarctic-Oscillation-like patterns can be found. Further, the SSTA composite maps indicate ENSO signals in near-concurrent lead times and Central-Pacific El Niño-like signals in long lead times.

*China – Yangtze River:*

- The case study in the Yangtze River shows that the JAS inflow can be predicted with strong forecasting skill. The optimum reliability is 0.833 (in the 3-month lead time), and the best dipole trace can explain about 46% of inflow variance (in the 8-month lead time).
- In the 8-month lead time, the South Indian Ocean and the ENSO region are identified as the major dipole components for the JAS inflow at the Three Gorges Dam. On the other hand, in the 4-month lead time, the East Coast of Africa and the Central Equatorial Pacific along with the continuously active South Indian Ocean are identified as tripole-like patterns.
- Composite maps regarding the JAS inflow at the Three Gorges Dam reveal several pronounced patterns related to the southwestern monsoon and Pacific jet

streams, but these patterns do not reverse from wet to dry years. In addition, the evolution of SSTA composites indicates that high (low) JAS inflow typically occurs in the decaying years of El Niño (La Niña) events.

*Central Africa – Congo River Basin:*

- The case study in the Congo River shows that the NDJ water level can be predicted with strong forecasting skill. The optimum reliability is 0.867, and the best dipole trace can explain about 52% of water-level variance in the 11- to 12-month lead times.
- Despite the strong forecasting skill, well-organized dipole components can hardly be found in the 11-month lead time. However, in the 1-month lead time, clear zonal-oriented dipoles composed of negative poles located at the South Atlantic and positive poles located at the Central Indian Ocean are identified for the NDJ water level at the Kinshasa station.
- Composite maps regarding the NDJ water level at the Kinshasa station illustrate that more pronounced patterns are far from the Congo River basin, and these patterns do not reverse from wet to dry years. Besides, the evolution of SSTA composites indicates that high NDJ water level typically occurs when ENSO experiences a transition from La Niña to El Niño phase, and vice versa.

## 7.2 Recommendations for Future Work

This dissertation can be extended in several ways:

- From an application standpoint, the dipole teleconnection method can be applied to hydro-climatic predictands of interests in other world regions potentially influenced by teleconnection patterns.
- To further assess the benefits of forecasting results obtained from the new method in other fields (e.g., water resources management), it is required to generate year-round, continuous forecasting traces that account for inter-seasonal correlations. One of the many approaches may be the combination of Bayesian statistics and historical analogue.
- The dipole method can also incorporate predictors other than SSTs. Such predictors could include additional upper-level pressure and wind fields. Some studies (e.g., Philippon et al., 2002; Hastenrath et al., 2004; Rajeevan et al., 2007) have already shown potential benefit, especially when the upper limit of predictability may result from the sole dependency on the SST field (Westra and Sharma, 2010). However, dealing with multi-predictor structures and temporal lags between these predictors may require further investigations.
- Hurricanes and tropical storms that exhibit different responses to SST forcing require a separate forecasting scheme. For example, hurricane-related statistics (e.g., number of landing hurricanes on a delineated spatial domain) can also be modeled through the dipole algorithm.
- Some of the hindcasting experiments have shown that wet and dry hydro-climatic events do not necessarily correspond to reversed large-scale patterns/mechanisms.

This suggests that an individual predictor-predictand relationship (based on different dipoles) could be developed pertaining to either wet or dry events.

- In addition to some dipole/SSTA patterns in certain lead times, it is assumed that the evolution of dipole/SSTA patterns should also contribute to local climatic conditions. One possible extension of the dipole algorithm is to design an add-on approach that accounts for the contribution of progressive dipole/SSTA variations.
- Ideally, the physical significance of identified dipole patterns can be verified by conducting GCM experiments that perturb SSTA in corresponding dipole areas.
- The predictability of local climatic conditions is not only governed by identified dipole patterns but also affected by the predefined spatial scale of predictands. Physically, the spatial scale of certain meteorological phenomena is region- and season-dependent, implying that the spatial scale of predictands should also be varied in a more sophisticated manner.
- Most of the statistical climate forecasting methods, including the dipole method, presume that both predictors and predictands are stationary so that analysis can be simplified. However, a climate regime shift caused by low-frequency oscillation (e.g., interdecadal modes) can also be operational. Thus, incorporating such climatic modes/patterns in the forecasting scheme would also be worthy of future investigations.

# APPENDIX

# CANONICAL CORRELATION ANALYSIS

An overview of canonical correlation analysis (CCA) used in Chapters 4 and 5 is presented. CCA, first proposed by Hotelling (1936), has a main concept that seeks the maximum correlation between two datasets through their linear transformation. Because CCA bears some resemblance to multiple regression, it can simply be converted into a forecasting framework with routinely updated predictors. Other advantages of CCA include concise derivations and straightforward interpretations.

## Canonical Correlations and Canonical Variables

Suppose there are two data vectors $X$ and $Y$:

$$X = \begin{bmatrix} x_1(t_1) & x_1(t_2) & \cdots & x_1(t_n) \\ x_2(t_1) & x_2(t_2) & \cdots & x_2(t_n) \\ \vdots & \vdots & \ddots & \vdots \\ x_I(t_1) & x_I(t_2) & \cdots & x_I(t_n) \end{bmatrix} \text{ and}$$

$$Y = \begin{bmatrix} y_1(t_1) & y_1(t_2) & \cdots & y_1(t_n) \\ y_2(t_1) & y_2(t_2) & \cdots & y_2(t_n) \\ \vdots & \vdots & \ddots & \vdots \\ y_J(t_1) & y_J(t_2) & \cdots & y_J(t_n) \end{bmatrix}, \tag{A.1}$$

where $(I, J)$ and $t_n$ are the number of observations in space and time. For the convenience of following interpretations, $X$ is referred to as the predictor variable (e.g., SSTA) and $Y$ as the predictand variable (e.g., precipitation). Instead of calculating the direct correlation between $X$ and $Y$ (which may be insignificant), it is desired to pursue

the maximum correlation between the "transformed" variables $a^T X$ and $b^T Y$, which can mathematically be formulated as

$$\begin{aligned}\rho &= \max_{a,b} corr\left(a^T X, b^T Y\right) \\ &= \max_{a,b} \frac{a^T \Sigma_{XY} b}{\sqrt{a^T \Sigma_{XX} a}\sqrt{b^T \Sigma_{YY} b}},\end{aligned} \tag{A.2}$$

where $\Sigma_{XX} = \mathrm{cov}(X, X)$, $\Sigma_{YY} = \mathrm{cov}(Y, Y)$, and $\Sigma_{XY} = \mathrm{cov}(X, Y)$. If the transformed variables, $v = a^T X$ and $w = b^T Y$, are subject to unit variance, that is, $Var(v) = a^T \Sigma_{XX} a = b^T \Sigma_{YY} b = Var(w) = 1$, $\rho$ is referred to as the maximum canonical correlation. In fact, the similar transformation can be repeated up to $M = \min(I, J)$ times in order to find other canonical correlations. These transformed variables are referred to as canonical variables:

$$\begin{aligned}&v_m = a_m^T X,\ m = 1, 2, \ldots, M, \text{ and} \\ &w_m = b_m^T Y,\ m = 1, 2, ..., M.\end{aligned} \tag{A.3}$$

The canonical correlations between each pair of transformed variables should be uncorrelated and can be arranged in descending order, namely

$$\begin{aligned}&corr\left(v_k, w_m\right) = \begin{cases}\rho_m, \text{ if } k = m, \\ 0, \text{ otherwise;}\end{cases} \text{ and} \\ &corr\left(v_1, w_1\right) \geq corr\left(v_2, w_2\right) \geq \cdots \geq corr\left(v_M, w_M\right) \geq 0.\end{aligned} \tag{A.4}$$

**Solving Procedure and Additional Properties**

Obtaining canonical correlations requires the solutions of vectors $a_m$ and $b_m$ (i.e., canonical vectors), which can be represented as an optimization problem equivalent to maximizing the numerator of Equation (A.1) subject to

$$a_m^T \Sigma_{XX} a_m = 1, \text{ and}$$
$$b_m^T \Sigma_{YY} b_m = 1. \tag{A.5}$$

Such problem can be solved by studying a Lagrange function defined as

$$\Lambda(\lambda, a_m, b_m) = a_m^T \Sigma_{XY} b_m - \frac{\lambda_X}{2}\left(a_m^T \Sigma_{XX} a_m - 1\right) - \frac{\lambda_Y}{2}\left(b_m^T \Sigma_{YY} b_m - 1\right), \tag{A.6}$$

and taking its derivatives with respect to $a_m$ and $b_m$:

$$\frac{\partial \Lambda}{\partial a_m} = \Sigma_{XY} b_m - \lambda_X \Sigma_{XX} a_m = 0,$$
$$\frac{\partial \Lambda}{\partial b_m} = \Sigma_{YX} a_m - \lambda_Y \Sigma_{YY} b_m = 0. \tag{A.7}$$

By combining these two equations into one by subtracting the second equation times $b_m$ from the first one times $a_m$, it yields

$$\lambda_Y b_m^T \Sigma_{YY} b_m - \lambda_X a_m^T \Sigma_{XX} a_m = 0, \tag{A.8}$$

which implies $\lambda = \lambda_X = \lambda_Y$. Thus, if $\Sigma_{YY}$ is invertible, Equation (A.7) can be rewritten as

$$b_m = \frac{\Sigma_{YY}^{-1} \Sigma_{YX} a_m}{\lambda}. \tag{A.9}$$

Substituting this term back to Equation (A.8) can produce

$$\Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} a_m = \lambda^2 \Sigma_{XX} a_m. \tag{A.10}$$

Since the covariance matrices should be symmetric and positive semi-definite, the Cholesky decomposition can be applied to $\Sigma_{XX}$, that is,

$$\Sigma_{XX} = \Sigma_{XX}^{1/2} \left(\Sigma_{XX}^{1/2}\right)^T, \tag{A.11}$$

and let $\left(\Sigma_{XX}^{1/2}\right)^T a_m = e_m$, Equation (A.10) may become

$$\left(\Sigma_{XX}^{-1/2}\right)\left(\Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX}\right)\left(\Sigma_{XX}^{-1/2}\right)^T e_m = \lambda^2 e_m, \tag{A.12}$$

which is a symmetric eigenproblem of the form $Ax = \lambda x$. Therefore, $e_m$ should be an eigenvector of

$$\left(\Sigma_{XX}^{-1/2}\right)\left(\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\right)\left(\Sigma_{XX}^{-1/2}\right). \tag{A.13}$$

Similarly, the other vector corresponding to $b_m, f_m$, should be an eigenvector of

$$\left(\Sigma_{YY}^{-1/2}\right)\left(\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\right)\left(\Sigma_{YY}^{-1/2}\right). \tag{A.14}$$

The solutions of canonical vectors are then simply written as

$$\begin{aligned} a_m &= \Sigma_{XX}^{-1/2} e_m, \text{ and} \\ b_m &= \Sigma_{YY}^{-1/2} f_m, \end{aligned} \tag{A.15}$$

where $m = 1, \ldots, M$. Further inspections can reveal that the associated canonical correlations would be the positive square roots of the $M$ nonzero eigenvalues,

$$\rho_m = \sqrt{\lambda_m}, \; m = 1, \ldots, M. \tag{A.16}$$

If CCA is applied to geophysical data, sometimes two additional correlations between the canonical and original variables are also calculated and may be helpful to discover any underlying physics. The first one is referred to as homogeneous correlations:

$$\begin{aligned} corr\left(v_m, X^T\right) &= a_m^T \Sigma_{XX} D_X^{-1}, \text{ and} \\ corr\left(w_m, Y^T\right) &= b_m^T \Sigma_{YY} D_Y^{-1}, \end{aligned} \tag{A.17}$$

where $D_X$ and $D_Y$ are diagonal matrices that contain standard deviations of $X$ and $Y$, and the second one is referred to as heterogeneous correlations:

$$\begin{aligned} corr\left(v_m, Y^T\right) &= a_m^T \Sigma_{XY} D_Y^{-1}, \text{ and} \\ corr\left(w_m, X^T\right) &= b_m^T \Sigma_{YX} D_X^{-1}. \end{aligned} \tag{A.18}$$

Most of time, the homogeneous and heterogeneous correlations derived from the leading canonical variables $(v_1,\ w_1)$ which exhibit the maximum canonical correlation would only be examined.

**Use CCA in Forecasting**

By using the idea of linear regression, the predictand canonical variables ($w_m$) can be related to the predictor canonical variables ($v_m$):

$$w_m = \hat{\beta}_{0,m} + \hat{\beta}_{1,m} v_m,\ m = 1,\ldots,M. \tag{A.19}$$

If the original variables $X$ and $Y$ are zero-mean, unit-variance vectors, simple derivations can reveal the intercept and slope values in Equation (A.19) as

$$\begin{aligned} \hat{\beta}_{0,m} &= 0, \text{ and} \\ \hat{\beta}_{1,m} &= \rho_m, \end{aligned} \tag{A.20}$$

for every $m = 1,\ldots,M$. As soon as the updated predictor canonical variables become available, the predictand canonical variables can be forecasted through the simple linear regression:

$$\hat{\mathbf{w}} = R_C \mathbf{v}, \tag{A.21}$$

where

$$R_C = \begin{bmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_M \end{bmatrix}, \tag{A.22}$$

and the canonical variables $(\hat{\mathbf{w}}$ and $\mathbf{v})$ are $(M \times 1)$ vectors. To transform back to the original predictand variable, it is simply to take the inverse of Equation (A.3):

$$\hat{Y} = \mathbf{B}^{-1}\mathbf{w}, \tag{A.23}$$

where the matrix $\mathbf{B}$ is filled with the corresponding $M$ canonical vectors in each row.

# REFERENCES


Adams, R. M., K. J. Bryant, B. A. Mccarl, D, M. Legler, J. J. O'Brien, A. Solow, and R. Weiher, 1995: Value of improved long-range weather information. *Contemporary Econ. Pol.*, **13**(3), 10–19.

Alexander, M. A., I. Bladé, M. Newman, J. R. Lanzante, N.-C., Lau, and J. D. Scott, 2002: The atmospheric bridge: The influence of ENSO teleconnections on air–sea interaction over the global oceans. *J. Climate*, **15**, 2205–2231.

Amarasekera, K. N., R. F. Lee, E. R. Williams, E. A. B. Eltahir, 1997: ENSO and the natural variability in the flow of tropical rivers. *J. Hydrol.*, **200**, 24–39.

Anderson, J., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, **9**, 1518–1530.

Anderson, J., H. van den Dool, A. Barnston, W. Chen, W. Stern, and J. Ploshay, 1999: Present-day capabilities of numerical and statistical models for atmospheric extratropical seasonal simulation and prediction. *Bull. Amer. Meteor. Soc.*, **80**, 1349–1361.

Ashok, K., S. K. Behera, S. A. Rao, H. Weng, and T. Yamagata, 2007: El Niño Modoki and its possible teleconnection. *J. Geophys. Res.*, **112**, C11007, doi:10.1029/2006JC003798.

Ashok, K., Z. Guan, and T. Yamagata, 2003: Influence of the Indian Ocean dipole on the Australian winter rainfall. *Geophys. Res. Lett.*, **30**, 1821, doi:10.1029/2003GL017926.

Balas, N., S. E. Nicholson, and D. Klotter, 2007: The relationship of rainfall variability in West Central Africa to sea-surface temperature fluctuations. *Int. J. Climatol.*, **27**, 1335–1349.

Barnston, A. G., 1994: Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J. Climate*, **7**, 1513–1564.

Barnston, A. G., and R. E. Livezey, 1987: Classification, seasonal and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.*, **115**, 1083–1126.

Barnston, A. G., and C. F. Ropelewski, 1992: Prediction of ENSO episodes using canonical correlation analysis. *J. Climate*, **5**, 1316–1345.

Barnston, A. G., and Y. X. He, 1996: Skill of canonical correlation analysis forecasts of 3-month mean surface climate in Hawaii and Alaska. *J. Climate*, **9**, 2579–2605.

Barnston, A. G., M. H. Glantz, and Y. X. He, 1999: Predictive skill of statistical and dynamical climate models in SST forecasts during the 1997-98 El Niño episode and the 1998 La Niña onset. *Bull. Amer. Meteor. Soc.*, **80**, 217–243.

Barnston, A. G., S. Li, S. J. Mason, D. G. Dewitt, L. Goddard, and X. Gong, 2010: Verification of the first 11 years of IRI's seasonal climate forecasts. *J. Appl. Meteorol.*, **49**, 493–520.

Barnett, T. P., 1995: Monte Carlo climate forecasting. *J. Climate*, **8**, 1005–1022.

Barnett, T. P., and R. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for United States surface air temperature determined by canonical correlation analysis. *Mon. Wea. Rev.*, **115**, 1825–1850.

Behera, S. K., J-J. Luo, S. Masson, P. Delecluse, S. Gualdi, A. Navarra, and T. Yamagata, 2005: Paramount impact of the Indian Ocean Dipole on the East African short rains: A GCM study. *J. Climate*, **18**, 4514–4530.

Bell, G. D., and J. E. Janowiak, 1995: Atmospheric circulation associated with the Midwest floods of 1993. *Bull. Amer. Meteor. Soc.*, **76**, 681–695.

Bengtsson, L., U. Schlese, E. Roeckner, M. Latif, T. P. Barnett, and N. Graham, 1993: A two-tiered approach to long-range climate forecasting. *Science*, **261**, 1026–1029.

Bierkens, M. F. P., and L. P. H. van Beek, 2009: Seasonal predictability of European discharge: NAO and hydrological response time. *J. Hydrometeor.*, **10**, 953–968.

Bjerknes, J., 1969: Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Rev.*, **97**, 163–172.

Black, E., J. Slingo, and K. R. Sperber, 2003: An observational study of the relationship between excessively strong short rains in coastal East Africa and Indian Ocean SST. *Mon. Wea. Rev.*, **131**, 74–94.

Block, P., and B. Rajagopalan, 2007: Interannual variability and ensemble forecast of Upper Blue Nile Basin Kiremt season precipitation. *J. Hydrometeor.*, **8**, 327–343.

Bove, M. C., J. B. Elsner, C. W. Landsea, X. Niu, and J. J. O'Brien, 1998: Effect of El Niño on U.S. landfalling hurricanes, revisited. *Bull. Amer. Meteor. Soc.*, **79**, 2477–2482.

Branković, Č., and T. Palmer, 1997: Atmospheric seasonal predictability and estimates of ensemble size. *Mon. Wea. Rev.*, **125**, 859–874.

Camberlin, P., and N. Philippon, 2002: The East African March–May rainy season: Associated atmospheric dynamics and predictability over the 1968–97 period. *J. Climate*, **15**, 1002–1019.

Camberlin, P., and R. E. Okoola, 2003: The onset and cessation of the "long rains" in eastern Africa and their interannual variability. *Theor. Appl. Climatol.*, **75**, 43–54.

Cane, M. A., S. E. Zebiak, and S. C. Dolan, 1986: Experimental forecasts of El Niño. *Nature*, **321**, 827–832.

Cane, M. A., G. Eshel, and R. W. Buckland, 1994: Forecasting Zimbabwean maize yield using eastern equatorial Pacific sea surface temperature. *Nature*, **370**, 204–205.

Cassou, C., L. Terray, J. W. Hurrell, and C. Deser, 2004: North Atlantic winter climate regimes: Spatial asymmetry, stationarity with time, and oceanic forcing. *J. Climate*, **17**, 1055–1068.

Cayan, D. R., K. T. Redmond, L. G. Riddle, 1999: ENSO and hydrologic extremes in the western United States. *J. Climate*, **12**, 2881–2893.

Chan, J. C., 1985: Tropical cyclone activity in the Northwest Pacific in relation to the El Niño/Southern Oscillation phenomenon. *Mon. Wea. Rev.*, **113**, 599–606.

Chan, J. C., J. S. Shi, and C. M. Lam, 1998: Seasonal forecasting of tropical cyclone activity over the western North Pacific and the South China Sea. *Wea. Forecasting*, **13**, 997–1004.

Chan, J. C., J. E. Shi, and K. S. Liu, 2001: Improvements in the seasonal forecasting of tropical cyclone activity over the western North Pacific. *Wea. Forecasting*, **16**, 491–498.

Chang, C.-P., Y. Zhang, and T. Li, 2000a: Interannual and interdecadal variations of the East Asian summer monsoon and tropical Pacific SSTs. Part I: Roles of the subtropical ridge. *J. Climate*, **13**, 4310–4325.

Chang, C.-P., Y. Zhang, and T. Li, 2000b: Interannual and interdecadal variations of the East Asian summer monsoon and tropical Pacific SSTs. Part II: Meridional structure of the monsoon. *J. Climate*, **13**, 4326–4340.

Chen, D., and C. Hellström, 1999: The influence of the North Atlantic Oscillation on the regional temperature variability in Sweden: spatial and temporal variations. *Tellus*, **51**A, 505–516.

Chiew, F. H. S., T. C. Piechota, J. A. Dracup, and T. A. McMahon, 1998: El Nino/Southern Oscillation and Australian rainfall, streamflow, and drought: Links and potential for forecasting. *J. Hydrol.*, **204**, 138–149.

Chu, P. S., and X. Zhao, 2007: A Bayesian regression approach for predicting seasonal tropical cyclone activity over the central North Pacific. *J. Climate*, **20**, 4002–4013.

Clark, C. O., P. J. Webster, and J. E. Cole, 2003: Interdecadal variability of the relationship between the Indian Ocean zonal mode and East African coastal rainfall anomalies. *J. Climate*, **16**, 548–554.

Colman, A., and M. Davey, 1999: Prediction of summer temperature, rainfall and pressure in Europe preceding winter North Atlantic Ocean temperature. *Int. J. Climatol.*, **19**, 513–536.

Cullen, H. M., A. Kaplan, P. A. Arkin, and P. B. Demenocal, 2002: Impact of the North Atlantic Oscillation on Middle Eastern climate and streamflow. *Clim. Change*, **55**, 315–338.

Croci-Maspoli, M., C. Schwierz, and H. C. Davis, 2007: Atmospheric blocking: space-time links to the NAO and PNA. *Clim. Dyn.*, **29**, 713–725.

Dai, A., and T. M. L. Wigley, 2000: Global patterns of ENSO-induced precipitation. *Geophys. Res. Lett.*, **27**, 1283–1286.

Daly, C., W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. Pasteris, 2002: A knowledge-based approach to the statistical mapping of climate. *Climate Research*, **22**, 99–113.

Diaz, H. F., and V. Markgraf, 2000: *El Niño and the Southern Oscillation: Multiscale Variability and Global and Regional Impacts*. Cambridge University Press, 512 pp.

Diro, G. T., D. I. F. Grimes, and E. Black, 2011a: Teleconnections between Ethiopian summer rainfall and sea surface temperature: part I – observation and modelling. *Clim. Dyn.*, **37**, 103–119, DOI 10.1007/s00382-010-0837-8.

Diro, G. T., D. I. F. Grimes, and E. Black, 2011b: Teleconnections between Ethiopian summer rainfall and sea surface temperature: part II – seasonal forecasting. *Clim. Dyn.*, **37**, 121–131, DOI: 10.1007/s00382-010-0896-x.

Drosdowsky, W., and L. E. Chambers, 2000: Note and correspondence: Near-global sea surface temperature anomalies as predictors of Australian seasonal rainfall. *J. Climate*, **14**, 1677–1687.

Efron, B., and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 456 pp.

Elsner, J. B., and C. P. Schmertmann, 1994: Assessing forecast skill through cross validation. *Wea. Forecasting*, **9**, 619–624.

Enfield, D. B., A. M. Mestas-Nuñez, and P. J. Trimble, 2001: The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S. *Geophys. Res. Lett.*, **28**, 2077–2080.

Fan, K., and H. Wang, 2009: A new approach to forecasting typhoon frequency over the western North Pacific. *Wea. Forecasting*, **24**, 974–986.

Feddersen, H., A. Navarra, and M. N. Ward, 1999: Reduction of model systematic error by statistical correction for dynamical seasonal predictions. *J. Climate*, **12**, 1974–1989.

Folland, C. K., J. A. Owen, M. N. Ward, and A. W. Colman, 1991: Prediction of seasonal rainfall in the Sahel region of Africa using empirical and dynamical methods. *J. Forecasting*, **10**, 2–56.

Fraedrich, K., and K. Müller, 1992: Climate anomalies in Europe associated with ENSO extremes. *Int. J. Climatol.*, **12**, 25–31.

Frankignoul, C., 1985: Sea surface temperature anomalies, planetary waves, and air-sea feedback in the middle latitudes. *Rev. Geophys.*, **23**(4), 357–390.

Gandin, L. S., and A. H. Murphy, 1992: Equitable skill scores for categorical forecasts. *Mon. Wea. Rev.*, **120**, 361–370.

Gates, W. L., 1992: AMIP: The Atmospheric Model Intercomparison project. *Bull. Amer. Meteor. Soc.*, **73**, 1962–1970.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 2003: *Bayesian Data Analysis*. Chapman and Hall/CRC, 696 pp.

Georgakakos, K. P., A. P. Georgakakos, and N. E. Graham, 1998: Assessment of benefit of climate forecasts of reservoir management in the GCIP region. *GEWEX News*, **8**(3), 5–7.

Georgakakos, K. P., N. H. Graham, F.-Y. Cheng, C. Spencer, E. Shamir, A. P. Georgakakos, H. Yao, and M. Kistenmacher, 2011a: Value of adaptive water resources management in Northern California under climatic variability and change: dynamic hydroclimatology. *J. Hydrol.*, in press, on line reference doi:10.1016/j.jhydrol.2011.04.032.

Georgakakos, A. P., H. Yao, M. Kistenmacher, K. P. Georgakakos, N. H. Graham, F.-Y. Cheng, C. Spencer, E. Shamir, 2011b: Value of adaptive water resources management in Northern California under climatic variability and change: reservoir management. *J. Hydrol.*, in press, on line reference doi:10.1016/j.jhydrol.2011.04.038.

Gerrity, J. P., 1992: Notes and correspondence: A note on Gandin and Murphy's equitable skill score. *Mon. Wea. Rev.*, **120**, 2709–2712.

Gershunov, A., and T. P. Barnett, 1998: ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: Observation and model results. *J. Climate*, **11**, 1575–1586.

Gershunov, A., and D. R. Cayan, 2003: Heavy daily precipitation frequency over the contiguous United States: Source of climatic variability and seasonal predictability. *J. Climate*, **16**, 2752–2765.

Gissila, T., E. Black, D. I. F. Grimes, and J. M. Slingo, 2004: Seasonal forecasting of the Ethiopian summer rains. *Int. J. Climatol.*, **24**, 1345–1358.

Glantz, M. H., 2001: *Currents of Change: Impacts of El Niño and La Niña on Climate and Society*. Cambridge University Press, 266 pp.

Glantz, M. H., R. W. Katz, and N. Nicholls, 1991: *Teleconnections Linking Worldwide Climate Anomalies*. Cambridge University Press, 545 pp.

Goddard, L., N. E. Graham, 1999: Importance of the Indian Ocean for simulating rainfall anomalies over eastern and southern Africa. *J. Geophys. Res.*, **104**(D16), 19099–19116.

Goddard, L., S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basher, and M. A. Cane, 2001: Current approaches to seasonal-to-interannual climate predictions. *Int. J. Climatol.*, **21**, 1111–1152.

Goddard, L., and S. J. Mason, 2002: Sensitivity of seasonal climate forecasts to persisted SST anomalies. *Clim. Dyn.*, **19**, 619–631.

Goddard, L., A. G. Barnston, and S. J. Mason, 2003: Evaluation of the IRI's "net assessment" seasonal climate forecasts 1997-2001. *Bull. Amer. Meteor. Soc.*, **84**, doi: 10.1175/BAMS-84-12-1761.

Goldenberg, S. B., C. W. Landsea, A. M. Mestas-Nuñez, and W. M. Gray, 2001: The recent increase in Atlantic hurricane activity: Causes and implications. *Science*, **293**, 474–479.

Gong, D., and S. Wang, 1999: Definition of Antarctic oscillation index. *Geophys. Res. Lett.*, **26**, 459–462.

Graham, R. J., A. D. L. Evans, K. R. Mylne, M. S. J. Harrison, and K. B. Robertson, 2000: An assessment of seasonal predictability using atmospheric general circulation models. *Q. J. R. Meteorol. Soc.*, **126**, 2211–2240.

Grantz, K., B. Rajagopalan, M. Clark, and E. Zagona, 2005: A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. *Water Resour. Res.*, **41**, W10410, doi:10.1029/2004WR003467.

Gray, W. M., 1984a: Atlantic seasonal hurricane frequency. Part I: El Niño and 30 mb Quasi-Biennial Oscillation influences. *Mon. Wea. Rev.*, **112**, 1649–1668.

Gray, W. M., 1984b: Atlantic seasonal hurricane frequency. Part II: Forecasting its variability. *Mon. Wea. Rev.*, **112**, 1669–1683.

Gray, W. M., C. W. Landsea, P. W. Mielke Jr., and K. J. Berry, 1992: Predicting Atlantic seasonal hurricane activity 6-11 months in advance. *Wea. Forecasting*, **7**, 440–455.

Gray, W. M., C. W. Landsea, P. W. Mielke Jr., and K. J. Berry, 1994: Predicting Atlantic seasonal tropical cyclone activity by 1 June. *Wea. Forecasting*, **9**, 103–115.

Hamlet, A. F., and D. P. Lettenmaier, 1999: Columbia river streamflow forecasting based on ENSO and PDO climate signals. *J. Water Resour. Plng. and Mgmt.*, **125**(6), 333–341.

Hansen, J. W., A. Challinor, A. Ines, T. Wheeler, and V. Moron, 2006: Translating climate forecasts into agricultural terms: advances and challenges. *Clim. Res.*, **33**, 27–41.

Harrison, D. E., and N. K. Larkin, 1998: El Niño-southern oscillation sea surface temperature and wind anomalies, 1946-1993. *Rev. Geophys.*, **36**(3), 353–399, doi:10.1029/98RG00715

Hastenrath, S., 1995: Recent advances in tropical climate prediction. *J. Climate*, **8**, 1519–1532.

Hastenrath, S., 2007: Circulation mechanisms of climate anomalies in East Africa and the equatorial Indian Ocean. *Dyn. Atmos. Oceans*, **43**, 25–35.

Hastenrath, S., D. Polzin, and P. Camberlin, 2004: Exploring the predictability of the 'short rains' at the coast of East Africa. *Int. J. Climatol.*, **24**, 1333–1343.

Hastenrath, S., D. Polzin, and C. Mutai, 2007: Diagnosing the 2005 drought in equatorial East Africa. *J. Climate*, **20**, 4628–4637.

Hastenrath, S., D. Polzin, and C. Mutai, 2011: Circulation Mechanisms of Kenya Rainfall Anomalies. *J. Climate*, **24**, 404–412.

Hirst, A. C., and S. Hastenrath, 1983: Diagnostics of hydrometeorological anomalies in the Zaire (Congo) Basin. *Q. J. R. Meteorol. Soc.*, **109**, 881–892.

Hotelling, H., 1936: Relations between two sets of variants. *Biometrika*, **28**, 321–377.

Hsieh, W. W., 2001: Nonlinear canonical correlation analysis of the tropical Pacific climate variability using a neural network approach. *J. Climate*, **14**, 2528–2539.

Hsieh, W. W., and B. Tang, 1998: Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bull. Amer. Meteor. Soc.*, **79**, 1855–1870.

Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science*, **269**, 676–679.

Hurrell, J. W., and van Loon, H., 1997: Decadal variations in climate associated with the North Atlantic oscillation. *Clim. Change*, **36**, 301–326.

Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck, 2003: *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*. American Geophysical Union, 279 pp.

Hurrell, J. W., and C. Deser, 2009: North Atlantic climate variability: The role of the North Atlantic Oscillation. *J. Mar. Syst.*, **78**, 28–41.

Indeje, M., F. H. M. Semazzi, and L. J. Ogallo, 2000: ENSO signals in East African rainfall seasons. *Int. J. Climatol.*, **20**, 19–46.

Janowiak, J. E., 1988: An investigation of interannual rainfall variability in Africa. *J. Climate*, **1**, 240–255.

Ji, M., D. W. Behringer, and A. Leetmaa, 1998: An improved coupled model for ENSO prediction and implications for ocean initialization. Part II: The coupled model. *Mon. Wea. Rev.*, **126**, 1022–1034.

Johansson, A., A. G. Barnston, S. Saha, H. M. and van den Dool, 1998: On the level of forecast skill in northern Europe. *J. Atmos. Sci.*, **55**, 103–127.

Johnson, M. A., and J. J. O'Brien, 1990: The northeast Pacific Ocean response to the 1982–1983 El Niño. *J. Geophys. Res.*, **95**, 7155–7166.

Jones, J. W., J. W. Hansen, F. S. Royce, and C. D. Messina, 2000: Potential benefits of climate forecasting to agriculture. *Agricult. Ecosys. Environ.*, **82**, 169–184.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.

Kanamitsu, M., A. Kumar, H.-M. H. Juang, J.-K. Schemm, W. Wang, F. Yang, S.-Y. Hong, P. Peng, W. Chen, S. Moorthi, and M. Ji, 2002: NCEP dynamical seasonal forecast system 2000. *Bull. Amer. Meteor. Soc.*, **83**, 1019–1037.

Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analysis of global sea surface temperature 1856–1991. *J. Geophys. Res.*, **103**, 18567–18589.

Kiladis, G. N., and H. F. Diaz, 1989: Global climatic anomalies associated with extremes in the Southern Oscillation. *J. Climate*, **2**, 1069–1090.

Kim, H. S., C. H. Ho, P. S. Chu, and J. H. Kim, 2010: Seasonal prediction of summertime tropical cyclone activity over the East China Sea using the least absolute deviation regression and the Poisson regression. *Int. J. Climatol.*, **30**, 210–219.

Klotzbach, P. J., 2008: Refinements to Atlantic basin seasonal hurricane prediction from 1 December. *J. Geophys. Res.*, **113**, D17109, doi:10.1029/2008JD010047.

Klotzbach, P. J., and W. M. Gray, 2004: Updated 6–11-month prediction of Atlantic basin seasonal hurricane activity. *Wea. Forecasting*, **19**, 917–934.

Korecha, D., and A. G., Barnston, 2007: Predictability of June to September rainfall in Ethiopia. *Mon. Wea. Rev.*, **135**, 628–650.

Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. Larow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, **13**, 4196–4216.

Kumar, A., A. Barnston, P. Peng, M. P. Hoerling, and L. Goddard, 2000: Changes in the spread of the variability of the seasonal mean atmospheric states associated with ENSO. *J. Climate*, **13**, 3139–3151.

Kung, E. C., and T. A. Sharif, 1980: Regression forecasting of the onset of the Indian summer monsoon with antecedent upper air conditions. *J. Appl. Meteorol.*, **19**, 370–380.

Kurtzman, D., and B. R. Scanlon, 2007: El Niño–Southern Oscillation and Pacific Decadal Oscillation impacts on precipitation in the southern and central United States: Evaluation of spatial distribution and predictions. *Water Resour. Res.*, **43**, W10427, doi:10.1029/2007WR005863.

Kwon, H. J., W. J. Lee, S. H. Won, and E. J. Cha, 2007: Statistical ensemble prediction of the tropical cyclone activity over the western North Pacific. *Geophys. Res. Lett.*, **34**, L24805, doi:10.1029/2007GL032308.

Lamb, P. J., and R. A. Peppler, 1987: North Atlantic Oscillation: Concept and an application. *Bull. Amer. Meteor. Soc.*, **68**, 1218–1225.

Landman, W. A., and S. J. Mason, 1999: Operational prediction of South African rainfall using canonical correlation analysis. *Int. J. Climatol.*, **19**, 1073–1090.

Landsea, C. W., and J. A. Knaff, 2000: How much skill was there in forecasting the very strong 1997–98 El Niño? *Bull. Amer. Meteor. Soc.*, **81**, 2107–2119.

Larkin, N. K., and D. E. Harrison, 2005a: On the definition of El Niño and associated seasonal average U.S. weather anomalies. *Geophys. Res. Lett.*, **32**, L13705, doi:10.1029/2005GL022738.

Larkin, N. K., and D. E. Harrison, 2005b: Global seasonal temperature and precipitation anomalies during El Niño autumn and winter. *Geophys. Res. Lett.*, **32**, L16705, doi:10.1029/2005GL022860.

Latif, M., T. P. Barnett, M. A. Cane, M. Flügel, N. E. Graham, H. von Storch, J.-S. Xu, and S. E. Zebiak, 1994: A review of ENSO prediction studies. *Clim. Dyn.*, **9**, 167–179.

Latif, M., D. Anderson, T. P. Barnett, M. A. Cane, R. Kleeman, A. Leetmaa, J. O'Brien, A. Rosati, and E. Schneider, 1998: A review of the predictability and prediction of ENSO. *J. Geophys. Res.*, **103**, 14375–14393.

Latif, M., D. Dommenget, M. Dima, and A. Grötzner, 1999: The role of Indian Ocean sea surface temperature in forcing East African rainfall anomalies during December–January 1997/98. *J. Climate*, **12**, 3497–3504.

Lau, N.-C., and M. J. Nath, 1996: The role of the "atmospheric bridge" in linking tropical Pacific ENSO events to extratropical SST anomalies. *J. Climate*, **9**, 2036–2057.

Lau, N.-C., and M. J. Nath, 2000: Impact of ENSO on the variability of the Asian-Australian monsoons as simulated in GCM experiments. *J. Climate*, **13**, 4287–4309.

Lavers, D., L. Luo, and E. F. Wood, 2009: A multiple model assessment of seasonal climate forecast skill for applications. *Geophys. Res. Lett.*, **36**, L23711, doi:10.1029/2009GL041365.

Leather, D. J., B. Yarnal, and M. A. Palecki, 1991: The Pacific/North American teleconnection pattern and United States climate. Part I: Regional temperature and precipitation associations. *J. Climate*, **4**, 517–528.

Leather, D. J., and M. A. Palecki, 1992: The Pacific/North American teleconnection pattern and United States climate. Part II: Temporal characteristics and index specification. *J. Climate*, **5**, 707–716.

Li, S., L. Goddard, and D. G. Dewitt, 2008: Predictive skill of AGCM seasonal climate forecasts subject to different SST prediction methodologies. *J. Climate*, **21**, 2169–2186.

Lima, C. H. R., U. Lall, T. Jebara, and A. G. Barnston, 2009: Statistical prediction of ENSO from subsurface sea surface temperature using a nonlinear dimensionality reduction. *J. Climate*, **22**, 4501–4519.

Lima, C. H. R., and U. Lall, 2010: Climate informed long term seasonal forecasts of hydroenergy inflow for Brazilian hydropower system. *J. Hydrol.*, **381**, 65–75.

Lo, F., M. C. Wheeler, H. Meinke, and A. Donald, 2007: Probabilistic forecasts of the onset of the north Australian wet season. *Mon. Wea. Rev.*, **135**, 3506–3520.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

Lu, M. M., P. S. Chu, and Y. C. Lin, 2010: Seasonal prediction of tropical cyclone activity near Taiwan using Bayesian multivariate regression method. *Wea. Forecasting*, **25**, 1780–1795.

Lyon, B., and D. G. DeWitt, 2012: A recent and abrupt decline in the East African long rains. *Geophys. Res. Lett.*, **39**, L02702, doi:10.1029/2011GL050337.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Amer. Meteor. Soc.*, **78**, 1069–1079.

Mantua, N. J., and S. R. Hare, 2002: The Pacific Decadal Oscillation. *J. Oceanogr.*, **58**, 35–44.

Markowski, G. R., and G. R. North, 2003: Climatic influence of sea surface temperature: Evidence of substantial precipitation correlation and predictability. *J. Hydrometeor.*, **4**, 856–877.

Mason, S. J., 1998: Seasonal forecasting of South African rainfall using a non-linear discriminant analysis model. *Int. J. Climatol.*, **18**, 147–164.

Mason, S. J., L. Goddard, N. E. Graham, E. Yulaeva, L. Sun, and P. A. Arkin, 1999: The IRI seasonal climate prediction system and the 1997/98 El Niño event. *Bull. Amer. Meteor. Soc.*, **80**, 1853–1873.

Mavromatis, T., S. S. Jagtap, and J. W. Jones, 2002: El Niño-Southern Oscillation effects on peanut yield and nitrogen leaching. *Clim. Res.*, **22**, 129–140.

McCabe, G. J., and M. D. Dettinger, 1999: Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States. *Int. J. Climatol.*, **19**, 1399–1410.

McCabe, G. J., M. A. Palecki, and J. L. Betancourt, 2004: Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. *Proc. Natl. Acad. Sci.*, **101**, 4136–4141.

Meza, F. J., J. W. Hansen, and D. Osgood, 2008: Economic value of seasonal climate forecasts for agriculture: Review of ex-ante assessments and recommendations for future research. *J. Appl. Meteorol.*, **47**, 1269–1286.

Michaelsen, J., 1987: Cross-validation in statistical climate forecast models. *J. Climate Appl. Meteor.*, **26**, 1589–1600.

Mjelde, J. W., H. S. J. Hill, and J. F. Griffiths, 1998: A review of current evidence on climate forecasts and their economic effects in agriculture. *Amer. J. Agr. Econ.*, **80**(5), 1089–1095.

Mo, K., and J. E. Schemm, 2008: Relationships between ENSO and drought over the southeastern United States. *Geophys. Res. Lett.*, **35**, L15701, doi:10.1029/2008GL034656.

Mo, R., and D. M. Straus, 2002: Statistical-dynamical seasonal prediction based on principal component regression of GCM ensemble integrations. *Mon. Wea. Rev.*, **130**, 2167–2187.

Moron, V., A. Navarra, M. N. Ward, and E. Roeckner, 1998: Skill and reproducibility of seasonal rainfall patterns in the tropics in ECHAM-4 GCM simulations with prescribed SST. *Clim. Dyn.*, **14**, 83–100.

Mutai, C. C., M. N. Ward, and A. W. Colman, 1998: Towards the prediction of the East Africa short rains based on sea-surface temperature – atmosphere coupling. *Int. J. Climatol.*, **18**, 975–997.

Neter, J., M. Kutner, C. Nachtsheim, and W. Wasserman, 1996: *Applied Linear Statistical Models*. McGraw-Hill/Irwin, 1408 pp.

Nicholls, N., 1979: A possible method for predicting seasonal tropical cyclone activity in the Australian region. *Mon. Wea. Rev.*, **107**, 1221–1224.

Nicholls, N., 1984: A system for predicting the onset of the north Australian wet-season. *J. Climatol.*, **4**, 425–435.

Nicholson, S. E., 1996: A review of climate dynamics and climate variability in Eastern Africa. In Johnson, T. C., and E. Odada (Eds.): *The Limnology, Climatology and Paleoclimatology of the East African Lakes*, Gordon & Breach, pp. 25–56.

Nicholson, S. E., and D. Entekhabi, 1986: The quasi-periodic behavior of rainfall variability in Africa and its relationship to the Southern Oscillation. *Arch. Met. Geoph. Biocl.*, Ser. A **34**, 311–348.

Nicholson, S. E., and J. Kim, 1997: The relationship of the El Niño-Southern Oscillation to African rainfall. *Int. J. Climatol.*, **17**, 117–135.

Ogallo, L. J., 1988: Relationships between seasonal rainfall in East Africa and Southern Oscillation. *J. Climatol.*, **8**, 31–43.

Palmer, T. N., and D. L. T. Anderson, 1994: The prospects for seasonal forecasting – A review paper. *Q. J. R. Meteorol. Soc.*, **120**, 755–793.

Palmer, T. N., Č. Branković, and D. S. Richardson, 2000: A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. *Q. J. R. Meteorol. Soc.*, **126**, 2013–2033.

Palmer, T. N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Délécluse, M. Déqué, E. diéz, F. J. Doblas-reyes, H. Feddersen, R. Graham, S. Gualdi, J.-F. Guérémy, R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar, E. Maisonnave, V. Marletto, A. P. Morse, B. Orfila, P. Rogel, J.-M. Terres, and M. C. Thomson, 2004: Development of a European multimodel ensemble system for

seasonal-to-interannual prediction (DEMETER). *Bull. Amer. Meteor. Soc.*, **85**, 853–872.

Philippon, N., P. Camberlin, and N. Fauchereau, 2002: Empirical predictability study of October-December East African rainfall. *Q. J. R. Meteorol. Soc.*, **128**, 2239–2256.

Pohl, B., and P. Camberlin, 2006: Influence of the Madden-Julian Oscillation of East African rainfall: II. March–May season extremes and interannual variability. *Q. J. R. Meteorol. Soc.*, **132**, 2541–2558.

Power, S., T. Casey, C. Folland, A. Colman, and V. Mehta, 1999: Inter-decadal modulation of the impact of ENSO on Australia. *Clim. Dyn.*, **15**, 319–324.

Rajagopalan, B., U. Lall, and S. Z. Zebiak, 2002: Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles. *Mon. Wea. Rev.*, **130**, 1792–1811.

Rajeevan, M., D. S. Pai, R. A. Kumar, and B. Lai, 2007: New statistical models for long-range forecasting of southwest monsoon rainfall over India. *Clim. Dyn.*, DOI 10.1007/s00382-006-0197-6.

Rasmusson, E. M., and T. H. Carpenter, 1983: The relationship between the eastern Pacific sea surface temperature and rainfall over India and Sri Lanka, *Mon. Wea. Rev.*, **111**, 354–384.

Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analysis using optimum interpolation. *J. Climate*, **7**, 929–948.

Robertson, A. W., U. Lall, S. Z. Zebiak, and L. Goddard, 2004: Improved combination of multiple atmospheric GCM ensembles for seasonal prediction. *Mon. Wea. Rev.*, **132**, 2732–2744.

Rogers, J. C., 1984: The association between the North Atlantic Oscillation and the Southern Oscillation in the Northern Hemisphere. *Mon. Wea. Rev.*, **112**, 1999–2015.

Rogers, J. C., 1988: Precipitation variability over the Caribbean and tropical Americas associated with the Southern Oscillation. *J. Climate*, **1**, 172–182.

Rogers, J. C., 1997: North Atlantic storm track variability and its association to North Atlantic Oscillation and climate variability of Northern Europe. *J. Climate*, **10**, 1635–1647.

Rogers, J. C. and H. van Loon, 1979: The seesaw in winter temperatures between Greenland and North Europe. Part II: Some oceanic and atmospheric effects in middle and high latitudes. *Mon. Wea. Rev.*, **107**, 509–519.

Ropelewski, C. F., and M. S. Halpert, 1986: North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114**, 2352–2362.

Ropelewski, C. F., and M. S. Halpert, 1987: Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Mon. Wea. Rev.*, **115**, 1606–1626.

Ropelewski, C. F., and M. S. Halpert, 1987: Precipitation patterns associated with the high index phase of the southern oscillation. *J. Climate*, **2**, 268–284.

Ropelewski, C. F., and M. S. Halpert, 1987: Quantifying Southern Oscillation – Precipitation relationships. *J. Climate*, **9**, 1043–1059.

Saha, S., S. Nadiga, C. Thiaw, J. Wang, W. Wang, Q. Zhang, H. van den Dool, H. L. Pan, S. Moorthi, D. Behringer, D. Stokes, M. Pena, S. Lord, G. White, W. Ebisuzaki, P. Peng, and P. Xie, 2006: The NCEP climate forecast system. *J. Climate*, **19**, 3483–3517.

Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, 1999: A dipole mode in the tropical Indian Ocean, *Nature*, **401**, 360–363.

Samba, G., D. Nganga, and M. Mpounza, 2008: Rainfall and temperature variations over Congo-Brazzaville between 1950 and 1998. *Theor. Appl. Climatol.*, **91**, 85–97.

Schlesinger, M. E., and N. Ramankutty, 1994: An oscillation in the global climate system of period 65–70 years. *Nature*, **367**, 723–726.

Seager, R., A. Tzanova, and J. Nakamura, 2009: Drought in the southeastern United States: Causes, variability over the last millennium, and the potential for future hydroclimate change. *J. Climate*, **22**, 5021–5045.

Segele, Z. T., and P. J. Lamb, 2005: Characterization and variability of Kiremt rainy season over Ethiopia. *Meteorol. Atmos. Phys.*, **89**, 153–180.

Segele, Z. T., P. J. Lamb, and L. M. Leslie, 2009: Large-scale atmospheric circulation and global sea surface temperature associations with Horn of Africa June–September rainfall. *Int. J. Climatol.*, **29**, 1075–1100.

Serreze, M. C., F. Carse, R. G. Barry, and J. C. Rogers, 1997: Iceland low cyclone activity: Climatological features, linkages with NAO, and relationship with recent changes in the Northern Hemisphere circulation. *J. Climate*, **10**, 453–464.

Shabbar, A., B. Bonsal, and M. Khandekar, 1997: Canadian precipitation patterns associated with the southern oscillation. *J. Climate*, **10**, 3016–3027.

Shabbar, A., J. P. Hung, and K. Higuchi, 2001: The relationship between the wintertime North Atlantic Oscillation and blocking episodes in the North Atlantic. *Int. J. Climatol.*, **21**, 355–369.

Sharma, A., 2000: Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 –A strategy for system predictor identification. *J. Hydrol.*, **239**, 232–239.

Sharma, A., K. C. Luk, I. Cordery, and U. Lall, 2000: Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 2 – Predictor identification of quarterly rainfall using ocean-atmosphere information. *J. Hydrol.*, **239**, 240–248.

Shepherd, M., T. Mote, J. Dowd, M. Roden, P. Knox, S. C. McCutcheon, and S. E. Nelson, 2011: An overview of synoptic and mesoscale factors contributing to the disastrous Atlanta flood of 2009. *Bull. Amer. Meteor. Soc.*, **92**, 861–870.

Shukla, J., J. Anderson, D. Baumhefner, C. Brankovic, Y. Chang, E. Kalnay, L. Marx, T. Palmer, D. Paolino, J. Ploshay, S. Schubert, D. Straus, M. Sharez, and J. Tribbia, 2000: Dynamical seasonal prediction. *Bull. Amer. Meteor. Soc.*, **81**, 2593–2606.

Sivillo, J. K., J. E. Ahlquist, and Z. Toth, 1997: An ensemble forecasting primer. *Wea. Forecasting*, **12**, 809–818.

Stockdale, T. N., D. L. T. Anderson, J. O. S. Alves, and M. A. Balmaseda, 1998: Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature*, **392**, 370–373.

Stone, M., 1974: Cross-validatory choice and assessment of statistical predictions. *J. R. Statist. Soc.*, **B36**, 111–147.

Tang, B., W. W. Hsieh, A. H. Monahan, and F. T. Tangang, 2000: Notes and correspondence: Skill comparisons between neural networks and canonical correlation analysis in predicting the equatorial Pacific sea surface temperatures. *J. Climate*, **13**, 287–293.

Tangang, F. T., B. Tang, A. H. Monahan, and W. W. Hsieh, 1998: Forecasting ENSO events: A neural network approach. *J. Climate*, **11**, 29–41.

Thompson, D. W. J., and J. M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: month-to-month variability. *J. Climate*, **13**, 1000–1016.

Thomson, M. C., F. J. Doblas-Reyes, S. J. Mason, R. Hagedorn, S. J. Connor, T. Phindela, A. P. Morse, and T. N. Palmer, 2006: Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*, **439**, 576–579.

Tippett, M. K., M. Barlow, and B. Lyon, 2003: Statistical correction of central southwest Asia winter precipitation simulations. *Int. J. Climatol.*, **23**, 1421–1433.

Tippett, M. K., L. Goddard, and A. G. Barnston, 2005: Notes and correspondence: Statistical correction of central southwest Asian winter precipitation. *J. Climate*, **18**, 1831–1843.

Todd, M. C., and R. Washington, 2004: Climate variability in central equatorial Africa: Influence from the Atlantic sector. *Geophys. Res. Lett.*, **26**, L23202, doi:10.1029/2004GL020975.

Trenberth, K. E., 1997: The definition of El Niño. *Bull. Amer. Meteor. Soc.*, **78**, 2771–2777.

Trigo, R. M., T. J. Osborn, and J. Corte-Real, 2002: The North Atlantic oscillation influence on Europe: Climate impacts and associated physical mechanisms. *Clim. Res.*, **20**, 9–17.

Trigo, R. M., D. Pozo-Vázquez, T. J. Osborn, Y. Castro-Díez, S. Gámiz-Fortis, and M. J. Esteban-Parra, 2004: North Atlantic Oscillation influence on precipitation, river flow and water resources in the Iberian Peninsula. *Int. J. Climatol.*, **24**, 925–944.

Ummenhofer, C. C., A. S. Gupta, M. H. England, and C. J. C. Reason, 2009: Contributions of Indian Ocean sea surface temperatures to enhanced East African rainfall. *J. Climate*, **22**, 993–1013.

Uvo, C. B., C. A. Repelli, S. E. Zebiak, and Y. Kushnir, 1998: The relationships between tropical Pacific and Atlantic SST and northeast Brazil monthly precipitation. *J. Climate*, **11**, 551–562.

van den Dool, H., 2007: *Empirical Methods in Short-Term Climate Prediction*. Oxford University Press, 215 pp.

van Loon, H., and J. C. Rogers, 1978: The seesaw in winter temperatures between Greenland and North Europe. Part I: General description. *Mon. Wea. Rev.*, **106**, 296–310.

van Oldenborgh, G. J., M. A. Balmaseda, L. Ferranti, T. N. Stockdale, and D. T. Anderson, 2005: Evaluation of atmospheric fields from the ECMWF seasonal forecasts over a 15-year period. *J. Climate*, **18**, 3250–3269.

Vinayachandran, P. N., N. H. Saji, and T. Yamagata, 1999: Response of the equatorial Indian Ocean to an anomalous wind event during 1994. *Geophys. Res. Lett.*, **26**, 1613–1616.

Vizy, E. K., and K. H. Cook, 2001: Mechanisms by which Gulf of Guinea and Eastern North Atlantic sea surface temperature anomalies can influence African rainfall. *J. Climate*, **14**, 795–821.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784–812.

Walker, G. T., 1923: Correlations in seasonal variations of weather. Part VIII. *Mem. Indian Meteor. Dept.*, **24**, 75–131.

Walker, G. T., 1924: Correlations in seasonal variations of weather. Part IX. *Mem. Indian Meteor. Dept.*, **24**, 275–332.

Walker, G. T., and E. W. Bliss, 1932: World Weather V. *Mem. Roy. Met. Soc.*, **4**, 53–84.

Wang, B., R. Wu, and X. Fu, 2000: Pacific-East Asian teleconnection: How does ENSO affect East Asian climate? *J. Climate*, **13**, 1517–1536.

Wang, H., R. Fu, A. Kumar, and W. Li, 2010: Intensification of summer rainfall variability in the southeastern United States during recent decades. *J. Hydrometeor.*, **11**, 1007–1018.

Ward, N. M., 1998: Diagnosis and short-lead prediction of summer rainfall in tropical North Africa at interannual and multidecadal timescales. *J. Climate*, **12**, 3167–3191.

Webster, P. J., and S. Yang, 1992: Monsoon and ENSO: Selectively interactive systems. *Q. J. R. Meteorol. Soc.*, **118**, 877–926.

Westra, S., and A. Sharma, 2010: An upper limit to seasonal rainfall predictability? *J. Climate*, **23**, 3332–3351.

Wettstein, J. J., and L. O. Mearns, 2002: The influence of the North Atlantic-Arctic Oscillation on mean, variance, and extremes of temperature in the northeastern United States and Canada. *J. Climate*, **15**, 3586–3600.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 648 pp.

Wilks, D. S., 2008: Improved statistical seasonal forecasts using extended training data. *Int. J. Climatol.*, **28**, 1589–1598.

William, A. P., and C. Funk, 2011: Westward extension of the warm pool leads to a westward extension of the Walker circulation, drying eastern Africa. *Clim. Dyn.*, **37**, 2417–2435, doi:10.1007/s00382-010-0984-y.

Wolter, K., and M. S. Timlin, 1993: Monitoring ENSO in COADS with a seasonally adjusted principal component index. *Proceedings of the 17th Climate Diagnostics Workshop, Norman, Oklahoma*, 52–57.

Wolter, K., and M. S. Timlin, 1998: Measuring the strength of ENSO – how does 1997/98 rank? *Weather*, **53**, 315–324.

Wood, A. W., and D. P. Lettenmaier, 2006: A test bed for new seasonal hydrologic forecasting approaches in the Western United States. *Bull. Amer. Meteor. Soc.*, **87**, 1699–1712.

Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenmaier, 2002: Long-range experimental hydrologic forecasting for the eastern United States. *J. Geophys. Res.*, **107**, No. D20, 4429, doi:10.1029/2001JD000659.

Wu, R., Z.-Z. Hu, and B. P. Kirtman, 2003: Evolution of ENSO-related rainfall anomalies in East Asia. *J. Climate*, **16**, 3741–3757.

Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558.

Xu, J.-S., and H. von Storch, 1990: Predicting the state of the Southern Oscillation using principal oscillation pattern analysis. *J. Climate*, **4**, 345–364.

Xu, K., C. Brown, H.-H. Kwon, U. Lall, J. Zhang, S. Hayashi, and Z. Chen, 2007: Climate teleconnections to Yangtze river seasonal streamflow at the Three Gorges Dam, China. *Int. J. Climatol.*, **27**, 771–780.

Yang, F., and K.-M. Lau, 2004: Trend and variability of China precipitation in spring and summer: linkage to sea-surface temperatures. *Int. J. Climatol.*, **24**, 1625–1644.

Yarnal, B., and H. F. Diaz, 1986: Relationships between extremes of the Southern Oscillation and the winter climate of the Anglo American Pacific coast. *Int. J. Climatol.*, **6**, 197–219.

Yeh, S. W., J. S. Kug, B. Dewitte, M. H. Kwon, B. P. Kirtman, and F. F. Jin, 2009: El Niño in a changing climate. *Nature*, **461**, 511–514.

Yu, Z. P., P. S. Chu, and T. Schroeder, 1997: Predictive skills of seasonal to annual rainfall variations in the US affiliated Pacific Islands: canonical correlation analysis and multivariate principal components regression approaches. *J. Climate*, **10**, 2586–2599.

Zhang, Y., J. M. Wallace, and D. S. Battisti, 1997: ENSO-like interdecadal variability. *J. Climate*, **10**, 1004–1020.

Zhang, Q., C.-Y. Xu, T. Jiang, and Y. Wu, 2007: Possible influence of ENSO on annual maximum streamflow of the Yangtze River, China. *J. Hydrol.*, **333**, 265–274.

Zwiers, F. W., and H. von Storch, 2004: On the role of statistics in climate research. *Int. J. Climatol.*, **24**, 665–680.